INSTYTUT BADAWCZY LEŚNICTWA
INSTITUT POLONAIS DES RECHERCHES FORESTIÉRES
Seria D. — Podręczniki — **Nr 4.**



DEZYDERY SZYMKIEWICZ

ZADANIA I METODY STATYSTYKI



WARSZAWA 1948

INSTYTUT BADAWCZY LEŚNICTWA
INSTITUT POLONAIS DES RECHERCHES FORESTIÉRES
Seria D. — Podręczniki — Nr 4.



DEZYDERY SZYMKIEWICZ

ZADANIA I METODY STATYSTYKI



WARSZAWA 1948

Papier 61 x 86, 70 g. kl. V. Nakład 4000 egz. Wykonano XI. 1948.

PAŃSTWOWE BYTOMSKIE ZAKŁADY GRAFICZNE Z. 126. - R 15180

WSTĘP

1. Zadania statystyki. Statystyka ma za zadanie opis stanów rzeczy na podstawie wielkiej ilości danych. Na przykład ustalono pewnego razu w Anglii na podziałce przyrządu mierniczego położenie jednej z kresek pod mikroskopem. Pomiar powtórzono 40 razy i otrzymano następujące wyniki[1]) w dziesięciotysięcznych częściach yarda.

3,68	2,81	5,48	3,28	3,11	4,65	3,76	3,78
4,76	3,27	4,59	3,22	2,75	4,08	2,64	3,98
4,15	4,51	2,98	3,91	5,08	4,43	4,21	5,21
2,95	3,43	5,23	4,43	6,35	3,26	4,45	2,28
3,78	2,48	3,95	4,10	4,49	4,84	2,66	4,18

Albo przy oznaczaniu liczby kwiatów w szczytowych koszykach bławatka (*Centaurea Cyanus*) autor tej książki stwierdził u 40 okazów wziętych na chybi trafi z pól w okolicach Zimnej Wody pod Lwowem następujące liczby kwiatów:

30	27	33	33	30	41	34	33	33	39
30	33	32	33	35	35	33	27	36	28
35	30	27	39	37	31	36	26	26	29
25	41	41	37	35	28	30	28	35	28

Takie dane mogą mieć zresztą charakter jakościowy, np. dotyczyć płci noworodków, barwy kwiatów itp. Jako przykład

[1]) Comptes Rendus. Tom 106, str. 784 (1888).

przytoczę dane co do płci 40 dzieci ochrzczonych w rzymsko-katolickiej parafii w Zimnej Wodzie w czasie od 3 września 1916 do 12 maja 1918 roku:

ż ż ż ż m ż ż m ż m ż ż m m m m m m m ż
ż ż ż ż ż ż m ż m ż m m ż ż ż m m ż m m

Ilość podobnych danych w ogromnej większości zagadnień jest nieograniczona — można przecież te same pomiary powtarzać tyle razy, ile się tylko spodoba i zbierać bławatki bez końca z coraz to innych pól albo ze specjalnych kultur. Jest na ogół niemożliwością brać pod uwagę wszystkie możliwe dane. Wypada zadawalać się ich częścią, starając się zresztą o to, by mieć ich jaknajwięcej. Mówi się przeto w statystyce o próbach takiej czy innej populacji, rozumiejąc pod próbą zbiór danych w rodzaju przytoczonych czterdziestek a pod populacją ogół istniejącego lub możliwego do uzyskania materiału faktycznego. Nazwa populacji pochodzi stąd, że statystyka zrodziła się z zagadnień społecznych i politycznych i dotyczyła początkowo głównie ludności. Dla uniknięcia nieporozumień, warto jest zaznaczyć, że populacja nie jest zbiorem jakichś ciał ani zjawisk, tylko zbiorem liczb określających cechy ciał lub zjawisk albo zbiorem jakościowych określeń tych cech.

Spotyka się zrzadka populacje ograniczone, naprzykład fakty dotyczące ludności danego kraju w danym czasie. W takich przypadkach można operować całością populacji, co jednak w praktyce jest zwykle niewykonalne. Jeżeli chodzi na przykład o określenie wzrostu ludności, niepodobna wymierzyć go u wszystkich ludzi danego kraju. Trzeba zadowolić się próbą, dotyczącą najwyżej kilkudziesięciu tysięcy osobników, których się bierze zwykle wśród poborowych.

Stwarza się w ten sposób konieczność sądzenia o populacji według próby, o całości na podstawie przypadkowej części jej, która przy nieograniczonym zakresie populacji jest znikomo mała, a nawet przy ograniczonym zakresie nigdy nie stanowi poważniejszej jej części. Charakter próby zależy

zawsze w pewnej mierze od przypadku. Pociąga to za sobą szerokie zastosowanie nauki o prawdopodobieństwie w statystyce. Zato też próby muszą mieć naprawdę przypadkowy charakter. Wszelkie dobieranie ich, nieświadome czy rozmyślne, prowadzi zawsze do fałszywych wniosków.

Trzeba jeszcze do tych ogólnych uwag o statystyce dodać wyjaśnienie wartości jej wyników. Pod tym względem zachodzą nieraz nieporozumienia. Chodzi o to, że wyniki badań statystycznych charakteryzują populacje, opisują je, ale nie dają nic ponadto. Szczególnie ciekawe nieporozumienie zachodzi pod tym względem odnośnie do dokładności pomiarów. Przyjmuje się powszechnie, że średnia arytmetyczna obliczona z wyników pomiarów zbliża się nieograniczenie do mierzonej wielkości w miarę zwiększenia ilości pomiarów. Otóż tak nie jest. Średnia obliczana z coraz liczniejszych pomiarów zbliża się istotnie do prawdziwej średniej charakteryzującej populację złożoną z wyników pomiarów wykonanych i niewykonanych, ale możliwych do wykonania. Nie wynika jednak z tego bynajmniej, by średnia tej populacji była identyczna z mierzoną wielkością. Jest to założenie niczem nieuzasadnione. Prawdziwa wartość mierzonej wielkości nie może być tą drogą ustalona i zdaje się nie da wogóle ustalić. Na ten temat prowadzili dyskusje ludzie tej miary co Laplace i Gauss ale bezskutecznie. Ciekawe uwagi w tej kwestii znajdzie czytelnik w rozprawie szwedzkiego astronoma Charliera pt. „Ueber das Fehlergesetz" (Meddelanden fran Lunds Astronomiska Observatorium N 24). Obszernie referuje odnośne dyskusje Czuber w książce pt. „Theorie der Beobachtungsfehler" (Lipsk 1891), w której jest podana źródłowa literatura.

W końcu trzeba omówić podział zajmującej nas nauki. Można w niej wyróżnić dwa działy. Jeden zajmuje się pojedyńczą populacją. Przedmiotem drugiego są związki zachodzące między populacjami, na przykład związek między wzrostem synów a wzrostem ojców, związek między barwą włosów a bar-

wą oczu. Związki te są ujmowane jako zależności funkcjonalne, nie przyczynowe. Dlatego też z dwóch populacji, pozostających ze sobą w związku, uważa się bez różnicy jedną lub drugą za zmienną niezależną. Na przykład, jak chodzi o związek między wzrostem synów a wzrostem ojców, rozpatruje się równocześnie zależność wzrostu synów od wzrostu ojców i zależność wzrostu ojców od wzrostu synów, jakkolwiek z pewnością w tym drugim przypadku nie zachodzi związek przyczynowy. Mówi się przeto w statystyce nie o zależności, lecz o współzależności między populacjami. W tym zagadnieniu statystycznym, tak samo jak we wszystkich innych, chodzi tylko o opis.

2. Materiały statystyczne. Mogą one mieć treść bardzo różną. Początkowo stosowano metody statystyczne tylko do zagadnień społecznych i politycznych (skład ludności, rozdział majątku, podatki itp.). Sama nazwa statystyki stąd pochodzi. Wywodzi się ona z łacińskiego wyrazu status, któremu w średniowieczu nadano sens równoznaczny z wyrazem państwo — stąd w języku angielskim państwo oznacza się wyrazem state, w niemieckim przez Staat, we francuskim przez état. Statystą dawniej nazywano męża stanu, co np. można widzieć w tekście „Hamleta" pochodzącym z roku 1602. Pierwszym, który użył wyrazu „statystyka" w obecnym jego znaczeniu, miał być Gotfryd Achenwall, profesor nauk politycznych w Getyndze, autor dzieła „Abriss der Staatswissenschaft der europäischen Reiche" (1749). Nie jest to pewne, niektórzy autorowie twierdzą, że nazwa ta była przed Achenwallem używana przez Schmeitzela[1]).

Następnie już w wieku XIX., za inicjatywą Belga Jakóba Quetelet'a (1796—1874) i zwłaszcza Anglików Franciszka Galtona (1822—1911) i Karola Pearsona (1857—1936) wprowadzono metody statystyczne w szerokiej mierze do nauk biologicznych. Wyrazem tej pracy angielskich badaczy jest czasopismo „Biometrika", wychodzące od ro-

[1]) Zob. André Liesse. La statistique. 5 wyd. Paryż 1933, str. 2.

ku 1902. *Biometrią* nazwano statystykę biologiczną. Jednocześnie metody statystyczne zastosowano w szerokiej mierze do zagadnień kinetycznej teorii materii i nawet do zagadnień dotyczących gwiazd stałych, nie licząc zastosowania tych metod do określenia błędów obserwacyjnych.

Te różnice w treści materiałów mają jednak małe znaczenie, — stosuje się w nich zawsze te same metody. Na przykład zagadnienie ilości płatków w kwiatach jaskrowatych jest opracowywane tą samą metodą, co częstość wypadków śmierci w armii od kopnięcia konia (zob. ust. 37). Natomiast poważne znaczenie *mają różnice w jednolitości materiałów.*

Można wogóle postawić na czele metodyki statystycznej wymaganie, by materiał był jednolity. Nie należy tego rozumieć dosłownie, bo w przyrodzie niema dwóch rzeczy i zdarzeń w zupełności do siebie podobnych. Nie tylko organizmy tej samej rasy nie są jednakowe, ale prawdopodobnie nawet atomy tego samego pierwiastka i zawarte w nich elektrony. A co dopiero mówić o zdarzeniach, jak pomiary, obserwacje i stany pogody. Chodzi tu o jednolitość względną. Trzeba natomiast zdawać sobie sprawę z jej stopnia, który zresztą może być różny, zależnie od charakteru zagadnień.

Uchybienia zasadzie jednolitości mogą doprowadzić do najbardziej niedorzecznych wniosków. Na przykład w okresie czasu około roku 1938 pewien odłam prasy francuskiej podniósł alarm z powodu rzekomo wielkiej przestępczości wśród cudzoziemców mieszkających we Francji, m. i. wśród Polaków. Wniosek ten był oparty na stosunku liczby więźniów danej narodowości do jej ogólnej liczebności we Francji. Okazało się jednak, że wielu cudzoziemców przebywa w więzieniu za przewinienia, których Francuz we Francji popełnić nie może, a mianowicie za nielegalny pobyt na francuskim terytorium. Po odrzuceniu tych przypadków, które nie mają nic wspólnego z przestępczością, okazało się, że stosunkowa liczba więźniów cudzoziemców niewiele się różni od odnośnej liczby więźniów Francuzów i o wy-

jątkowej przestępczości wśród cudzoziemców we Francji nie może być mowy.

Zasada jednolitości materiału statystycznego nie pozwala m. i. na łączenie w jedną populację wyników obserwacyj wykonanych przez różne osoby, gdyż każdy człowiek ma swoiste skłonności do popełniania błędów. Nie pozwala ona traktować łącznie dochodowość inwestycji w różnych częściach kraju z uwagi na różne warunki ekonomiczne. Nie pozwala na dokładniejsze porównanie obserwacyj meteorologicznych wykonanych w różnych miejscach, jeżeli pochodzą one z różnych okresów itd.

Największe bodaj znaczenie posiada omawiana zasada dla nauk biologicznych, a to z powodu istnienia w obrębie nieomal każdego gatunku, rozumianego w zwykłym tego słowa znaczenia, licznych nieraz gatunków elementarnych, zwanych odmianami, rasami, czystymi liniami itp. Są one nieraz do siebie bardzo podobne i różnią się tylko cechami ilościowymi. Materiał wtedy wydaje się jednolitym, co w istocie nie ma miejsca Taki przypadek zachodzi np. u fasoli. Stwierdził to Johannsen[1]) mierząc i ważąc osobno nasiona pochodzące od różnych jej osobników. Wynik tych pomiarów był następujący: można osobniki tej rośliny podzielić na grupy takie, że w obrębie każdej grupy średnia waga i wymiary nasion będą jednakowe, natomiast różne w różnych grupach. Te średnie wartości utrzymują się bez zmiany w dalszych pokoleniach, wobec czego każda taka grupa stanowi odrębny elementarny gatunek, nazwany przez Johannsena czystą linią. Pomieszane ze sobą osobniki różnych czystych linij fasoli nie dają się rozpoznać przez ważenie i pomiary nasion, bo odnośne wielkości wykazują odchylenia od średniej tego samego rzędu wielkości co różnica między średnimi, charakteryzującymi dane linie. Naprzykład czysta linia *B* ma średnią wagę nasion 520 mg. zaś linia *O* — 310 mg, ale jednocześnie w obu tych liniach są nasiona o wadze 250—520 mg.

---

[1]) W. Johannsen. Ueber Erblichkeit in Populationen und in reinen Linien — Jena (1903).

U fasoli wyróżnienie gatunków elementarnych było stosunkowo łatwe, gdyż można było otrzymać bez trudu potomstwo każdego osobnika oddzielnie bez udziału innych osobników, a to skutkiem samozapylenia właściwego tej roślinie. Gorzej jest, jeżeli roślina rozmnaża się przez krzyżowanie, jako to zwykle bywa. Wtedy para osobników, biorąca udział w rozmnażaniu, może należeć do różnych gatunków elementarnych Skutkiem tego gatunek będzie mieszaniną nie tylko gatunków elementarnych, ale jeszcze i mieszańców między nimi. Otrzymanie jednolitego materiału statystycznego staje się wtedy daleko trudniejsze.

Materiał biologiczny może być niejednolity jeszcze z innego powodu. W różnych warunkach bytowania ilościowe cechy organizmów zwykle wykształcają się różnie, na przykład wzrost ludzi wyraźnie maleje przy niedostatecznym odżywianiu. Najsilniejsze zmiany tego rodzaju wykazują rośliny. Szczególnie jaskrawym przykładem jest mak, którego liczne znamiona są nieraz przedmiotem rozważań statystycznych. Jest ich najczęściej 13, ale to tylko w dobrych warunkach wegetacji. Niekorzystne warunki pomniejszają ich ilość i to tak dalece, że w krańcowych wypadkach tworzą się kwiaty z dwoma tylko znamionami. Jeżeli zatem chcemy stwierdzić, jaki jest wymiar cechy właściwy danemu gatunkowi, trzeba brać osobniki, które wyrosły w jednakowych warunkach otoczenia. Otrzymany wynik będzie przytem ważny tylko dla danych warunków. Natomiast materiał pochodzący z różnego otoczenia nie będzie jednolity i da wyniki chwiejne.

Jeżeli chodzi o ustalenie cech gatunkowych metodami statystycznymi, trzeba następnie mieć na uwadze, że każda liczba w odnośnym materiale powinna odnosić się do innego osobnika. Jest to szczególnie ważne przy badaniu roślin, u których widzi się często narządy pozornie jednakowe, np. kwiaty. Dla uproszczenia pracy badacze są skłonni traktować równorzędnie cechy takich narządów bez względu na ich przynależność do tych czy innych osobników. Na przykład przy zliczaniach znamion w kwiatach maku brano wszystkie kwiaty.

Tymczasem kwiaty wykazują wyraźne różnice zależnie od położenia ich na roślinie. I tak autor tej książki otrzymał dla liczby znamion maku następujące średnie u 100 osobników wziętych z pewnej kultury Ogrodu Botanicznego w Dublanach:

kwiaty szczytowe	12.80
kwiaty górnej gałęzi	12.14
kwiaty drugiej gałęzi	12.61
kwiaty trzeciej gałęzi	13.05

Zatem gdyby się wzięło wszystkie kwiaty maku bez względu na ich położenie na roślinie, miałoby się materiał niejednolity. Trzeba brać tylko kwiaty szczytowe albo z pewnej określonej gałęzi, po jednemu z każdego osobnika.

Na tym jeszcze nie koniec. Oddzielne rośliny, które się widzi w przyrodzie albo w kulturach, nie zawsze stanowią odrębne osobniki — mogą one być odrębnie żyjącymi częściami tego samego osobnika. Nie dotyczy to maku, który jest jednoroczną rośliną. Każdy jego okaz wyrasta z osobnego nasienia i jest odrębnym osobnikiem. Jeżeli jednak weźmiemy trwałe rośliny rozmnażające się wegetatywnie, jak poziomki albo ziemniaki, możemy mieć mnóstwo pozornie odrębnych roślin, będących tylko częściami tego samego osobnika — całe pole ziemniaków może stanowić jeden osobnik! Takie pozornie odrębne osobniki noszą nazwę k l o n ó w. Otóż różnice między klonami, pochodzącymi od tego samego osobnika, są znacznie mniejsze niż między naprawdę odrębnymi osobnikami. Przeto zaliczanie do tej samej populacji cech klonów, pochodzących od różnych osobników, da materiał niejednolity. W dziedzinie zoologii ma się takie trudności tylko u niższych zwierząt.

3. Metody statystyki. Mnogość materiału, którym operuje statystyka, powoduje to, że musi ona posługiwać się liczbami nawet wtedy, kiedy materiał opracowywany ma charakter jakościowy. W tym ostatnim przypadku ustala się ilość danych dla tej lub innej cechy. Na przykład w trzecim przykładzie ust. 1 mamy 18 chłopców i 22 dziewczyny. Z konieczności tedy m e t o d y s t a t y s t y k i m a j ą c h a r a k t e r m a t e m a-

tyczny. Statystyka jednak nie jest działem matematyki.

Jest to kwestia zasadnicza nie tylko dla statystyki, lecz dla wszystkich nauk zajmujących się badaniem rzeczywistości, czy to w dziedzinie przyrody, czy w zakresie nauk społecznych Zachodzi bowiem zasadnicza różnica między wspomnianymi naukami, które można nazwać „realistycznymi", a naukami matematycznymi. Mianowicie, podczas gdy w matematyce założenia mogą być dowolne, w naukach realistycznych muszą one wytrzymać próbę sprawdzania. Sprawdzanie polega na porównywaniu wniosków wyprowadzonych z założeń z obserwowanymi faktami. Wymagana jest zgodność jednych z drugimi, podczas gdy w matematyce wystarcza logiczna konsekwencja przy wyprowadzaniu wniosków.

W stosowaniu metod matematycznych do zagadnień statystycznych spotykamy się z niemałą trudnością, wynikającą z charakteru danych statystycznych. Mianowicie liczby materiałów, którymi operuje statystyka są zawsze całkowite. Jest to oczywiste w przypadkach, kiedy materiał otrzymuje się przez zliczanie — kwiatów w kwiatostanach, chłopców wśród dzieci itp. Mniej oczywiste są przypadki, w których dokonuje się pomiarów. Ma się wtedy liczby pozornie ułamkowe, jak te, które zostały przytoczone w pierwszym przykładzie ust. 1. Są to jednak ułamki z tym samym mianownikiem, a więc właściwie liczby całkowite, podające ilość jednostek miary użytej przy pomiarach. We wspomnianym przykładzie taką jednostką jest milionowa część yarda. Dopiero opracowanie materiałów daje liczby naprawdę ułamkowe, kiedy się oblicza z różną dokładnością średnie arytmetyczne, prawdopodobieństwa błędów i inne wielkości charakteryzujące populacje.

Niema przeto ciągłości w materiałach statystycznych, podczas gdy matematyka operuje przeważnie zmiennymi ciągłymi. Ten brak ciągłości jest spowodowany w pewnej mierze szerokim rozpowszechnieniem nieciągłości w przyrodzie, daleko

szerszym niż się to wydaje na pierwszy rzut oka. Pozornie ciągła materia okazała się zbiorem odrębnych cząsteczek — drobin, drobiny — zbiorem atomów, atomy — zbiorem protonów, neutronów i elektronów. Ostatnio „zatomizowano" nawet energię, przynajmniej promienistą, na kwanty i fotony. Skutkiem tego „atomizowania" przyrody pomiary coraz częściej są zastępowane przez zliczania. Naprzykład dla charakterystyki pierwiastków chemicznych używa się obecnie numerów zamiast ciężarów atomowych. Te numery są to liczby elementarnych dodatnich ładunków w jądrach atomowych. I tak wodór, ze swoim ciężarem atomowym 1.008, jest określany jako pierwiastek numer pierwszy i to wraz ze swoimi dwoma odmianami (izotopami), z którym jeden ma atomy dwa razy, a drugi trzy razy cięższe. Chlor, posiadający ciężar atomowy 34.45, stał się ze swoimi dwoma izotopami o ciężarach atomowych 35 i 37 pierwiastkiem numer siedemnasty itd. Trzeba przy tym jednak mieć na uwadze, że gdyby nawet ciągłość panowała w przyrodzie, to nieuniknione użycie jednostek w pomiarach spowodowałoby zawsze rozbicie jej na wielokrotności użytych jednostek.

Matematyczne metody statystyki są bardzo zawiłe i wymagają dobrej znajomości rachunku nieskończonościowego i nauki o prawdopodobieństwie. Chcąc umożliwić korzystanie z tej książki możliwie szerokim kołom czytelników, nie będę mógł podać pełnego uzasadnienia omawianych metod. Natomiast będę się starał na przykładach z możliwie różnych dziedzin wyjaśnić ich znaczenie dla należytego zobrazowania rzeczywistości.

Opracowanie metod statystycznych jest głównie zasługą angielskich badaczy, w szczególności Pearsona, Fishera oraz Gosseta, który niewiadomo dlaczego ukrywał się pod pseudonimem Studenta. Następnie dużo wnieśli do tej dziedziny pracy naukowej Skandynawowie, zwłaszcza szwedzki astronom Charlier, oraz Rosjanie: Czebyszew, Czuprow. Poważny jest udział polski, zwłaszcza dzięki pracom ś.p. Władysława Bortkiewicza.

Co się tyczy nauki o prawdopodobieństwie, która odgrywa wielką rolę w metodach statystycznych, to dzieje jej sięgają bardziej odległych czasów niż dzieje statystyki, czasów Błażeja Pascala (1623—1662) i Jakóba Bernoulli'ego (1654—1705).

W formie współczesnej została ona ugruntowana przez Laplace'a (1749—1827) w dziele pt. „Théorie analytique des probabilités" ogłoszonym w ostatecznej formie jako 3 wydanie w r. 1820. Paryska Akademia Nauk wydała je ponownie w roku 1886.

4. Literatura i pomoce. W języku polskim mamy trzy podręczniki statystyki. Dwa są oryginalne: J. Czekanowskiego „Zarys metod statystycznych w zastosowaniu do antropologii" (Warszawa 1913) oraz S. Moszczeńskiego „Metody statystyczne w zastosowaniu do organizacji gospodarstw rolniczych, ogrodniczych i leśnych" (Warszawa 1924). Obie te książki; mimo swojego specjalnego zadania, dają dobry przegląd metod statystyki.

Trzecim podręcznikiem w języku polskim jest tłumaczenie książki angielskiego badacza Yule'a „Wstęp do teorii statystyki" (Warszawa 1921). Jest pisany przystępnie, ale nieco zawile.

Pożyteczne są wreszcie artykuły A. Łomnickiego, drukowane w „Kosmosie, Seria B" z lat 1928 i 1930 pt. „Zagadnienia statystyki matematycznej".

W literaturze obcojęzycznej najbardziej przystępny wykład metod statystycznych znajdzie czytelnik w książce duńskiego botanika W. Johannsena „Elemente der exacten Erblichkeitslehre" (Jena 3 wyd. 1926).

Bardzo cenne są dalej dwa przystępnie pisane podręczniki. Jeden C. V. L. Charliera „Vorlesungen über die Grundzüge der matematischen Statistik" (Lund 1920). Drugi rosyjski B. U. Romanowskiego „Elementarnyj kurs matematiczeskoj statistiki" (Moskwa — Leningrad 1939).

Wymienione powyżej książki nie wymagają znajomości rachunku nieskończonościowego. Dla czytelników mających na-

leżyte przygotowanie matematyczne można polecić dwa podręczniki. Szczególnie cenna jest książka szwedzkiego statystyka Oskara N. Andersona „Einführung in die mathematische Statistik" (Wiedeń 1935). Pisana jest z werwą i humorem, co kontrastuje osobliwie z wielkim aparatem matematycznym. Z konieczności trudna w czytaniu, przedstawia jasno omawiane zagadnienia. Autor sprzeciwia się stanowczo traktowaniu statystyki, nawet tzw. matematycznej, jako działu matematyki, co znalazło swój wyraz na wstępie w podaniu motta z dzieł Czuprowa: „Statistik spielende Mathematiker können nur durch mathematisch ausgerüstete Statistiker überwunden werden". Ciekawa jest osoba Andersona — Szwed podający się za za ucznia Rosjanina Czuprowa i Polaka Bortkiewicza, uwzględniający skrupulatnie dorobek polski, podający tytuły prac polskich z dokładną pisownią, zatrudniony w Sofji jako dyrektor Instytutu Statystycznego do badań gospodarczych na Uniwersytecie.

Nie mniej cenna jest książka R. A. Fishera, najwybitniejszego obecnie angielskiego statystyka, pt. „Statistical methods for research workers" (Edymburg — Londyn, 6 wyd. 1936). Niestety, pisana stylem telegraficznym, jest zbiorem łamigłówek, pomimo licznych przykładów.

Źródłowa literatura teoretycznej statystyki jest silnie rozproszona. Są jednak dwa czasopisma specjalnie jej poświęcone: wspomniana powyżej w ust. 2 angielska „Biometrika" oraz międzynarodowe czasopismo „Metron" wychodzące we Włoszech.

Co się tyczy nauki o prawdopodobieństwie, mamy w języku polskim jeden tylko podręcznik W. Gosiewskiego: „Zasady rachunku prawdopodobieństwa" (Warszawa 1906). Z licznych książek w obcych językach można polecić poza podstawowym dziełem Laplace'a zbiorowe dzieło wychodzące pod redakcją Emila Borela od r. 1925 pt. „Traité de calcul des probabilités". Teoria statystyki jest nim objęta również.

Do wykonania rachunków bardzo pożyteczne są maszyny do rachowania (arytmometry). Między innymi bardzo wygodna

jest szwedzka maszyna systemu Odhner. Niemiecka Brunsviga jest nieznaczną odmianą tej szwedzkiej maszyny o niecą gorszym wykonaniu. Niestety, te maszyny są drogie i nie każdemu dostępne.

Można jednak i bez nich prowadzić badania statystyczne. Trzeba w tym celu zaopatrzyć się w tabele iloczynów, np. „Rechentafeln" Zimmermanna (Berlin 1918), które dają iloczyny liczb trzycyfrowych przez dwucyfrowe i w znacznej mierze zastępują maszynę do rachowania, o ile chodzi o mnożenie. Dają one także kwadraty, sześciany i pierwiastki kwadratowe liczb trzycyfrowych. Jeżeli jednak chodzi o dodawanie i odejmowanie, nic nie może zastąpić maszyny do rachowania. W braku jej trzeba uzbroić się w cierpliwość i z rezygnacją przyjąć nieuniknioną stratę czasu.

Tablice logarytmiczne są oczywiście niezbędne, ale te można zawsze mieć. Wystarczą 5-cyfrowe.



CZĘŚĆ PIERWSZA

STATYSTYKA POJEDYNCZYCH POPULACJI

ROZDZIAŁ I.

SZEREGI ROZDZIELCZE

5. Układanie szeregu rozdzielczego. Pierwszym krokiem w opracowaniu każdego jako tako licznego materiału statystycznego, mającego postać zobrazowaną przykładami ust. 1, jest ułożenie *szeregu rozdzielczego.* Poszczególne dane w takim materiale noszą nazwę *wariantów.* Takie warianty są częściowo jednakowe, częściowo różne. *Trzeba policzyć ile razy powtarza się każdy rodzaj wariantów i zestawić ich częstości. Takie zestawienie nazywa się szeregiem rozdzielczym.*

Najprościej przedstawia się omawiane zadanie, jeżeli materiał statystyczny składa się z danych jakościowych. Ma się wtedy do czynienia z niewielką ilością rodzajów wariantów I tak w trzecim przykładzie ust. 1 mamy tylko chłopców i dziewczynki, pierwszych jest 18, drugich 22. Bardziej zawiłe jest to zadanie w przypadkach materiału ilościowego — jest wtedy więcej różnych wariantów. A więc w przykładzie 2 ustępu 1 mamy dla liczby kwiatów w szczytowych koszykach bławatka 15 różnych wartości od 25—41. Napiszmy w jednym wierszu te liczby we wzrastającym porządku ich wielkości, nie pomijając tych, które nie są wcale reprezentowane. Pod nimi umieścimy ich częstości, przyczem częstość nieobecnych wartości oznaczymy przez zero.

25	26	27	28	29	30	31	32	33	34
1	2	3	4	1	5	1	1	7	1

(1)

35	36	37	38	39	40	41
5	2	2	0	2	0	3

Ułożyliśmy w ten sposób szereg rozdzielczy dla danego materiału.

Wartości wariantów w takim szeregu można uważać za wartości pewnego rodzaju zmiennej niezależnej, tzw. z m i e n n e j e w e n t u a l n e j. Różni się ona od zmiennych niezależnych rozpatrywanych w matematyce tym, że przybiera tylko wartości całkowite, a więc jest nieciągła. Odnośne częstości możemy uważać za pewnego rodzaju funkcje takiej zmiennej niezależnej. Będą one zawsze dodatnie i całkowite albo równe zeru, zatem ta funkcja będzie się odznaczała również brakiem ciągłości.

Forma otrzymanego szeregu (1) jest prowizoryczna i w dalszym ciągu trzeba będzie ją ulepszyć, ale już teraz uzyskujemy pewne informacje o badanej populacji. Przede wszystkim widzimy, w jakich granicach są zawarte liczbowe wartości wariantów, inaczej mówiąc określamy z a k r e s z m i e n n o ś c i wariantów. Następnie stwierdzamy, że pośrednie wartości wariantów wystąpiły na ogół w większej ilości, są częstsze, natomiast krańcowe — zarówno najmniejsze jak i największe — są w mniejszej ilości, są rzadsze.

Rzucającą się w oczy wadą rozpatrywanej formy szeregu są gwałtowne przeskoki w ciągu częstości, spowodowane przez przypadkowość próby, np. 34 występuje raz jeden, podczas gdy sąsiednie wartości wariantów — 33 i 35 — 7 i 5 razy. Tak jest zawsze, kiedy wielkość próby jest mała w porównaniu z wielkością wariantów. Lepiej to wypada, jeżeli warianty są mniejsze. Naprzykład jeżeli weźmiemy liczbę płatków u przylaszczki *(Anemone Hepatica)*, która to liczba waha się w granicach od 6 do 11, to próba tej wielkości, jaka była wzięta u bławatka, już może dać szereg wyrównany. I tak naprzykład próba:

7	7	7	7	6	8	9	7	7	6
8	7	10	6	7	6	8	8	6	6
7	6	6	6	6	6	6	6	6	8
7	7	6	6	7	8	6	6	8	6

daje szereg rozdzielczy bez przeskoków w częstościach

6	7	8	9	10
19	12	7	1	1

Zwiększenie próby wyrównuje stopniowo omawiane nieprawidłowości szeregu rozdzielczego. Trzeba jednak dużego materiału, by osiągnąć ten cel bodaj w części, jeżeli wartości wariantów są znaczne. Na przykład jeżeli do czterdziestkowej próby bławatka dodamy 260 wariantów z tego samego materiału, by mieć próbę z 300, otrzymamy szereg następujący:

21	22	23	24	25	26	27	28	29	30	
1	1	3	2	4	5	7	19	12	12	
31	32	33	34	35	36	37	38	39	40	
19	30	22	18	23	15	15	19	16	6	
										(2)
41	42	43	44	45	46	47	48	49	50	
15	13	5	4	0	4	1	0	3	1	
51	52	53	54	55	56	57	58	59	60	
3	0	0	1	0	0	0	0	0	0	
61	62									
0	1									

W nowej formie szereg wykazuje w środkowej części mniejsze przeskoki. Zupełne ich usunięcie jest niemożliwe nawet w wielotysięcznych próbach — prawie zawsze zostają one, co prawda tylko na końcach szeregu.

Przytoczony powyżej większy materiał wykazuje rozszerzenie zakresu zmienności — wartości wariantów są teraz w granicach od 21—62, zamiast 25 i 41. Dalsze powiększenie próby dałoby prawdopodobnie dalsze rozszerzenie, zresztą tylko do pewnej granicy. Widzimy zatem, że zakres zmienności zależy nietylko od materiału, lecz od wielkości próby.

PRZEDMOWA.

Sprawa rozpowszechnienia, a zwłaszcza pogłębienia znajomości metod statystyki matematycznej jest obecnie bardzo aktualnym zagadnieniem nauki polskiej. Dotychczas nie posiadamy bowiem w kraju ośrodka gromadzącego teoretyków-specjalistów pracujących twórczo w tym dziale wiedzy, jakkolwiek w dziedzinie jej zastosowań praktycznych możemy się wykazać poważnym dorobkiem. Metody statystyki matematycznej zdobyły sobie bowiem u nas prawa obywatelstwa nie tylko w różnych działach nauk biologicznych, a przede wszystkiem w antropologii, lecz sięgnęły też i do dziedziny nauk humanistycznych. Zarówno w etnologii jak i w językoznastwie, inicjatywa wyszła przy tym ze strony nauki polskiej, rozwijającej i unowocześniającej prawie zupełnie zapomniane poczynania znakomitego angielskiego etnologa E. B. TYLOR'a.

Ześrodkowanie się na zastosowaniach angielskiej statystyki matematycznej i na oryginalnym rozwinięciu aparatu indukcyjnego umożliwiło stworzenie polskiej szkoły antropologicznej, przeciwstawiającej się angielskiemu kierunkowi biometrycznemu swoimi mendelistycznymi założeniami, i pozwoliło na ilościowe uzasadnienie polskiej syntezy slawistycznej. Ponadto przeniesienie tych osiągnięć metodologicznych do dziedziny botaniki, stworzyło podstawy obiektywnego ustosunkowania się do zagadnień fitosocjologii, a w dziedzinie językoznastwa umożliwiło nawet uporanie się z zagadnieniem systematyki języków australijskich.

Mimo tak poważnych osiągnięć zastosowań nowoczesnej statystyki matematycznej w różnych dziedzinach wiedzy w bieżącym stuleciu i przy posiadaniu tak wybitnego teoretyka, jakim w poprzedniej generacji był śp. Władysław BORTKIEWICZ,

a obecnie jest Jerzy Spława-NEYMAN, profesor Uniwersytetu Berkeley w Kalifornii, nie posiadamy dotychczas uniwersyteckiego kompendium tej dziedziny wiedzy.

Śp. Dezydery SZYMKIEWICZ był tak wyjątkowym biologiem, że posiadane wykształcenie matematyczne pozwalało mu na podjęcie trudu opracowania tego rodzaju podręcznika, jednak był on tak przeciążony swoimi zajęciami zawodowymi, że sprawa poszła w odwłokę. Dopiero częściowa przymusowa bezczynność spowodowana przez wypadki wojenne i w tej dziedzinie zaktualizowała pracę nad podręcznikami uniwersyteckimi. W tym dopiero okresie został napisany ten na znacznie skromniejszą miarę zakrojony podręcznik. Był on planowany jako zarys elementarny, umożliwiający osiągnięcie orjentacji ogólnej i praktyczne stosowanie metod statystycznych. Stanowi on wynik prac przygotowawczych do oddawna planowanego kompendium, nieukończonego wskutek przedwczesnej śmierci wielce zasłużonego uczonego.

W Poznaniu, dnia 23. VIII. 1948 r.

Jan Czekanowski

Obecność przeskoków w częstościach powoduje, że szeregi rozdzielcze poddaje się przekształceniom. Mianowicie łączy się wartości wariantów w grupy równej wielkości, tzw. klasy, i oblicza się częstości dla tych klas przez sumowanie częstości zawartych w nich wartości wariantów. Ilość łączonych wartości należy wybrać o tyle dużą, żeby częstości obliczone dla klas zmieniały się stopniowo. Jeżeli w naszym materiale bławatka utworzymy klasy po dwie wartości, tego celu nie osiągniemy — przeskoki wprawdzie złagodnieją, ale zostaną. Widoczne to jest z następującego zestawienia:

21—22	23—24	25—26	27—28	29—30	31—32
2	5	9	26	24	49
33—34	35—36	37—38	39—40	41—42	43—44
40	38	34	22	28	9
45—46	47—48	49—50	51—52	53—54	55—56
4	1	4	3	1	0
57—58	59—60	61—62			
0	0	1			

(3)

Trzeba wobec tego użyć podziałów klasowych większych. Spróbujmy łączyć wartości wariantów po trzy. Możemy to uczynić w trojaki sposób, poczynając od wartości 19, 20 albo 21. W ogóle jest tyle sposobów łączenia wariantów w klasy, ile jest wartości w klasie. Dla bławatka wobec tego będziemy mieli do wyboru trzy różne formy szeregu (4), (5) i (6).

19—21	22—24	25—27	28—30	31—33	34—36
1	6	16	43	71	56
37—39	40—42	43—45	46—48	49—51	52—54
50	34	9	5	7	1
55—57	58—60	61—63			
0	0	1			

(4)

20—22	23—25	26—28	29—31	32—34	35—37
2	9	31	43	70	53
38—40	41—43	44—46	47—49	50—52	53—55
41	33	8	4	4	1
56—58	59—61	62—64			
0	0	1			

(5)

21—23	24—26	27—29	30—32	33—35	36—38
5	11	38	51	63	49
39—41	42—44	45—47	48—50	51—53	54—56
37	22	5	4	3	1
57—59	60—62				
0	1				

(6)

Najlepiej wyrównaną formą szeregu jest trzecia (6). We wszystkich trzech bardzo nieprzyjemnie zaznacza się skrajny wariant 62, odległy od wszystkich innych. Takie skrajne warianty czasem się odrzuca, czasem przeciwnie nadaje się im szczególne znaczenie. Kwestia ta będzie osobno omówiona w ust. 32. Wogóle w wyborze przedziałów klasowych należy kierować się tym, by częstości zmieniały się stopniowo. Poza tym dobrze jest zwrócić uwagę na to, by warianty o ile możności były skupione w środkach klas.

Szerokość klas nie powinna być zbyt duża, bo wtedy szereg wypadnie za krótki. Najdogodniejszą jest liczba klas między 15 a 25. Nasz szereg jest właściwie trochę za krótki, bo wprawdzie w pierwszych dwóch formach ma 15 wyrazów (w trzeciej 14), ale ostatni wyraz jest oddzielony od reszty dwiema (w trzeciej formie jedną) pustymi klasami.

Dalej ważną rzeczą przy opracowaniu szeregów rozdzielczych jest ustalenie *granic klas* i ich *wartości środ-*

kowych. Granicami klasy nie są jej krańcowe wartości wariantów. Na przykład dla klasy 31—33 ani 31 nie jest granicą, ani 33. *Granice klas wytycza się pośrodku między ostatnią wartością jednej klasy a pierwszą następnej*. W ten sposób granicami klasy 31—33 będą liczby $30\frac{1}{2}$ i $33\frac{1}{2}$. Będą to fikcyjne wartości zmiennej ewentualnej, żaden bowiem wariant nie ma ułamkowych wartości. Jest to jednak konieczne dla porównywania szeregów rozdzielczych z ciągłymi funkcjami, co będzie rozpatrywane w rozdziałach IV i VI.

Wyznaczanie środkowej wartości klasy, nazywanej także po prostu środkiem klasy, nie przedstawia żadnych trudności, jeżeli klasy zawierają nieparzystą ilość wartości wariantów. Tak jest z formami (4), (5) i (6) naszego szeregu, gdzie na przykład dla klasy 31—33 środkową wartością będzie 32. Zadanie jest bardziej zawiłe, jeżeli ilość wartości wariantów w klasach jest parzysta, jak w formie (3). Wtedy trzeba przyjąć jako wartość środkową fikcyjną wartość zmiennej ewentualnej po środku klasowego podziału. I tak w formie (3) dla klasy 31—32 wartością środkową będzie liczba $31\frac{1}{2}$.

Wartości środkowe są niezbędne dla rachunkowego opracowywania szeregów rozdzielczych. Wtedy, jak to zobaczymy w następnym rozdziale, służą one do wyznaczania *numerów klas*. *Te numery oblicza się, dzieląc wartości środkowe przez szerokość klas, to znaczy przez liczbę wartości wariantów w klasie*. Otrzymuje się liczby całkowite, albo mieszane, np. w formie (5) otrzymamy numery 7, 8, 9, . . . , w formie zaś (4) $6\frac{2}{3}$, $7\frac{2}{3}$, $8\frac{2}{3}$, . . . Liczby mieszane jako numery wyglądają trochę groteskowo, ale to nic nie szkodzi. Chodzi bowiem o to, że przez wprowadzenie ich można opracowywać szeregi tak, jakgdyby warianty nie były zgrupowane i jakgdyby w każdej klasie miały jednakową wartość, równą numerowi danej klasy. *Numery w ten sposób będą niczym innym jak osobliwą formą zmiennej ewentualnej*. Ze zwykłymi zmiennymi ewentualnymi mają one tę właściwość wspólną, że

w ciągach przez nie złożonych sąsiednie wyrazy różnią się o jedynkę, tak samo jak w ciągach wartości zwykłych zmiennych ewentualnych.

Jednocześnie z wprowadzeniem numerów klas trzeba zmienić liczbowe wartości granic, dzieląc je także przez szerokość klas. A więc granice dla szeregu formy (5) będą $6^1/_2$, $7^1/_2$, $8^1/_2$, ..., dla formy zaś (4) $6^1/_6$, $7^1/_6$, $8^1/_6$, ... Znowu otrzymujemy ciągi, których wyrazy różnią się o jedynkę.

Jest często rzeczą korzystną podawać dla wartości wariantów *częstości względne*, które otrzymuje się dzieląc częstości przez liczbę wariantów. W związku z tym dla uniknięcia nieporozumień, częstości nazywa się nieraz *częstościami bezwzględnymi*. Częstości względne są ułamkami, których suma powinna być równa jedności. W praktyce różni się ona zwykle cokolwiek od jedności z powodu odrzucania dalszych dziesiętnych.

Takie przeliczenia są pożyteczne między innymi przy porównywaniu podobnych materiałów, zwłaszcza przy porównywaniu różnych prób tej samej populacji. Naprzykład dla liczby kwiatów w szczytowych koszykach bławatka można przytoczyć oprócz podanego powyżej materiału z pól Zimnej Wody materiał pochodzący z pewnej kultury Ogrodu Botanicznego w Dublanach (tabela 5,1) a liczący 200 wariantów. W tej tabeli ciągi wartości zmiennej ewentualnej i ich częstości są wypisane w pionowych szeregach a nie poziomych, jak dotąd. Jest to zwykle wygodniejsze.

Zgrupujemy te dane w klasy po trzy, poczynając od wartości 19 jak w zestawieniu (4). Przeliczając otrzymane częstości na częstości względne i zestawiając je z częstościami względnymi z materiału z Zimnej Wody, otrzymamy następujące porównanie (tab. 5,2).

Bardzo pożyteczne jest graficzne przedstawianie szeregów rozdzielczych. Są na to dwa sposoby. W pierwszym z nich na osi odciętych oznacza się wartości zmiennej ewentualnej albo dla materiałów grupowanych numery klas.

Tabela 5,1

Liczby kwiatów w szczytowych koszykach bławatka z kultury Nr 149 Ogrodu Botanicznego w Dublanach

Liczby kwiatów	Ich częstości	Liczby kwiatów	Ich częstości	Liczby kwiatów	Ich częstości
19	1	30	11	41	4
20	3	31	14	42	2
21	1	32	19	45	2
22	1	33	14	51	1
23	4	34	13		
24	2	35	14		
25	6	36	11		
26	5	37	1		
27	12	38	11		
28	17	39	5		
29	23	40	3		

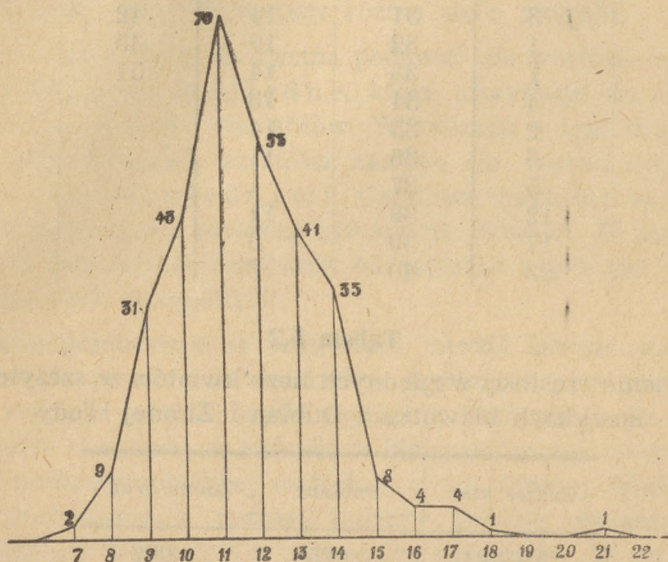
Tabela 5,2

Zestawienie częstości względnych liczb kwiatów w szczytowych koszykach bławatka z Dublan i Zimnej Wody

Zakres klas	Dublany	Zimna Wod
19—21	0,025	0,003
22—24	0,035	0,020
25—27	0,115	0,053
28—30	0,255	0,143
31—33	0,235	0,237
34—36	0,190	0,187
37—39	0,085	0,167
40—42	0,045	0,113
43—45	0,010	0,030
46—48	0,000	0,017
49—51	0,005	0,023
52—54	0,000	0,003
55—57	0,000	0,000
58—60	0,000	0,000
61—63	0,000	0,003
Sumy	1,000	0,999

Z tych punktów rysuje się rzędne o długościach proporcjonalnych do odnośnych częstości. Łącząc wierzchołki rzędnych, otrzymuje się linię łamaną (ryc. 1).

Dla materiału bławatka wznosi się ona i po osiągnięciu pewnego szczytu, odpowiadającego najczęstszej wartości, opada. W innych przypadkach może ta linia mieć inny kształt, np.



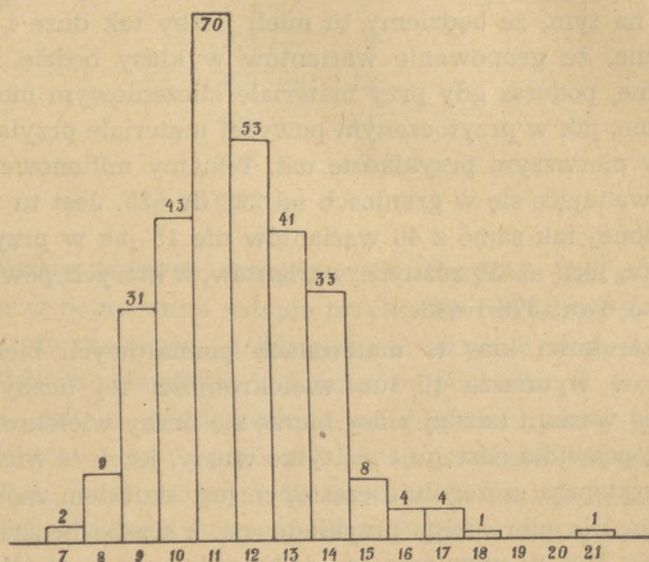
Ryc. 1. — Wykres zmienności dla liczby kwiatów w szczytowych koszykach bławatka z Zimnej Wody. Według drugiego sposobu grupowania materiału w klasy. Liczby u dołu oznaczają numery klas, liczby u góry — częstości wariantów w odnośnych klasach.

posiadać kilka wierzchołków itp. Nazywa się ją krzywą zmienności, jakkolwiek nie jest to krzywa.

Daleko ważniejszą jest druga forma omawianych wykresów. Na osi odciętych zaznaczamy teraz granice klas. Robimy to nawet wtedy, kiedy warianty nie są grupowane. Wówczas te granice wytyczamy po środku między wartościami wariantów. Na zaznaczonych w ten sposób odcinkach o długościach rów-

nych jedności budujemy prostokąty o wysokościach równych odnośnym częstościom. Otrzymamy „schodkowaną" figurę o formie w ogólnych zarysach podobnej do „krzywej" pierwszego sposobu (ryc. 2).

Takie figury noszą nazwę „histogramów" według terminu wprowadzonego przez Pearsona.



Ryc. 2. — Histogram dla tego samego materiału statystycznego co w rys. 1.

Histogramy są mało estetyczne. Ich wartość ujawnia się dopiero przy porównywaniu szeregów rozdzielczych z ciągłymi funkcjami, ale już teraz warto jest zaznaczyć, że pola prostokątów w histogramach są równe odnośnym częstościom, gdyż podstawy prostokątów są równe jedności. Suma tych pól równa się liczebności wariantów, w razie zaś użycia częstości względnych, równe są jedności.

Skale odciętych i rzędnych w obu rodzajach wykresów mogą być różne, gdyż chodzi tu o różne rodzaje wielkości, od-

cięte bowiem przedstawiają wartości zmiennej ewentualnej lub numery klas, rzędne zaś ich częstości.

Dotychczasowe wywody o szeregach rozdzielczych dotyczyły materiałów otrzymywanych przez zliczenia. W przypadku materiału pomiarowego opracowanie jego przedstawia się tak samo, gdyż ułamki figurujące w takich materiałach mają ten sam mianownik i są właściwie liczbami całkowitymi. Różnica polega na tym, że będziemy tu mieli liczby tak duże i przeto tak różne, że grupowanie wariantów w klasy będzie zawsze konieczne, podczas gdy przy materiale zliczeniowym może być zbyteczne, jak w przytoczonym powyżej materiale przylaszczki. I tak w pierwszym przykładzie ust. 1 mamy milionowe części yarda, wahające się w granicach od 228 do 635. Jest tu w próbie złożonej tak samo z 40 wariantów nie 15 jak w przypadku bławatka, lecz aż 38 różnych wariantów, z których powtarzają się tylko dwa: 378 i 443.

Szerokości klas w materiałach pomiarowych bierze się zwykle w wymiarze 10 lub wielokrotności tej liczby. Jako pierwszy wariant każdej klasy bierze się liczby wielokrotne 10. Od tego prawidła odstępuje się tylko wtedy, jeżeli te wielokrotności występują szczególnie często, co jest skutkiem zaokrąglania liczb. Dla pierwszego przykładu ust. 1 trzeba będzie wziąć szerokość klas w wymiarze stu jednostek pomiarowych i szereg będzie miał formę:

200—299	300—399	400—499	500—599	600—699
8	13	14	4	1

Trzeba przytem pamiętać, że jednostką pomiarową dla zmiennej ewentualnej jest w nim milionowa część yarda.

W tej formie szereg jest całkowicie wyrównany. Wadą jego jest to, że ma za mało wyrazów, co jest nieuniknione wobec szczupłości materiału. Zarówno granice klas, jak i wartości środkowe będą reprezentowane przez fikcyjne wartości zmiennej ewentualnej. Naprzykład dla klasy 400—499 granicami

będą liczby 399½ i 499½, środkiem zaś liczba 449,5 (nie 450!). Numerami klas będą liczby mieszane, naprzykład dla przytoczonej klasy będzie to liczba 4.495.

Dla skrócenia można klasy oznaczać przez same tylko początkowe wartości wariantów w klasach z dodaniem kreski. Naprzykład omawiany szereg pomiarowy można napisać w formie:

200 — 8
300 — 13
400 — 14
500 — 4
600 — 1

6. Formy prostych szeregów rozdzielczych. Już z przytoczonych w poprzednim ustępie przykładów widoczne są dwie różne formy szeregów rozdzielczych: dwuboczna, jak w przykładach bławatka i pomiarów podziałki, i jednoboczna, jak u przylaszczki. Jest jeszcze trzecia bardzo rzadka forma — szeregu wklęsłego. Inne jeszcze formy powstają, jeżeli materiał składa się z dwóch lub kilku różnych materiałów zestawionych razem. Będzie o tym mowa osobno w ust. 7.

Przyjrzyjmy się formie obu przytoczonych dwubocznych szeregów. Widzimy, że w środkowej ich części jedna z klas ma większą częstość, niż inne. Reprezentuje ona t. zw. modalną wartość wariantów, bardzo ważną dla charakterystyki materiałów statystycznych. Porównując ze sobą częstości klas położonych na jednakowej odległości po obu stronach klasy modalnej, widzimy, że są one nierówne: w przypadku bławatka większe są częstości wariantów większych od wartości modalnej, w przypadku pomiarów podziałki przeciwnie, większa jest częstość wariantów mniejszych. Oba szeregi wykazują zatem wyraźną asymetrię, inaczej mówiąc skośność.

Można powiedzieć, że w pierwszym z tych szeregów jest ona dodatnia, w drugim ujemna.

Tak jest zazwyczaj. Szeregi symetryczne są rzadkie. Klasycznym ich przykładem są szeregi rozdzielcze wzrostu ludzi. Najlepszym znanym autorowi materiałem tego rodzaju jest zestawienie, ogłoszone przez Lindersa[1]) (tab. 6,1). Dotyczy ono 46981 mężczyzn w wieku między 20 a 22 lat z armii i floty szwedzkiej. Wartości wzrostu są podane w przedziałach centymetrowych. Trzeba je wobec tego traktować tak, jak gdyby były wykonane w milimetrach.

Przykładem symetrycznej krzywej może służyć także otrzymany przez autora szereg rozdzielczy dla liczby kwiatów w koszykach sałatnicy (*Aposeris foetida*) (tab. 6,2) zebranej w lesie Suchowoli pod Lwowem. Z każdego okazu był wzięty jeden koszyk. Koszyki u tej rośliny są pojedyńcze, ale łodyg kwiatowych wyrasta u tego samego okazu zwykle większa ilość. Warto jest zaznaczyć, że tak samo w tym przykładzie, jak i poprzednim, są pewne odchylenia od symetrii — bez tego nigdy się nie obchodzi.

W przeciwieństwie do liczby kwiatów w koszykach ilość wewnętrznych listków okrywy u tej samej rośliny dała szereg z wyraźną symetrią dodatnią (por. drugą część tejże tabeli 6,2). Nie jest to bynajmniej fakt odosobniony, że różne cechy tego samego organizmu dają szeregi rozdzielcze o różnym charakterze. I tak podczas gdy wzrost ludzi układa się w szeregi symetryczne, ich waga daje szeregi asymetryczne. Jako przykład przytaczam za Yule'm dane angielskie. Wagę notowano w okrągłych funtach (tab. 6,3). Asymetria jest tu silna i wybitnie dodatnia.

Odnośnie do listków okrywy u sałatnicy warto jeszcze do-

---

[1]) F. J. Linders. Contributions to the knowledge of the stature and its variation within different social strata in Sweden. — Geografiska Annaler. 1930.

dać jedną ważną uwagę. Wyróżnia się ten materiał poza asymetrią jeszcze jedną ważną cechą, mianowicie tym, że modalna wartość ma częstość znacznie większą od sąsiednch.

Tabela 6,1

Wzrost 46981 Szwedów.

Wzrost w mm	Częstość	Wzrost w mm	Częstość
1510—	8	1740—	2788
1520—	7	1750—	2732
1530—	6	1760—	2332
1540—	13	1770—	2161
1550—	43	1780—	1726
1560—	69	1790—	1532
1570—	110	1800—	1153
1580—	179	1810—	926
1590—	292	1820—	719
1600—	450	1830—	535
1610—	607	1840—	385
1620—	851	1850—	277
1630—	1121	1860—	199
1640—	1359	1870—	143
1650—	1672	1880—	80
1660—	2146	1890—	56
1630—	2337	1910—	33
1690—	2865	1920—	17
1700—	3031	1930—	13
1710—	3115	1940—	9
1720—	3084	1950—	4
1730—	3093	1960—	3
		1970—	1

Tabela 6,2

Sałatnica (*Aposeris foetida*)

Ilość kwiatów w koszykach	Liczby osobników z takimi ilościami kwiatów	Ilości listków okrywy	Liczby osobników z takimi okrywami
2	11		
7	12	6	1
16	13	7	10
29	14	8	42
42	15	9	234
45	16	10	10
54	17	11	1
46	18		
43	19		
36	20		
19	21		
15	22		
1	23		
2	24		
1	25		
Ogółem	359	Ogółem	300

Tego rodzaju szeregi rozdzielcze są częste u roślin, m. i. kiedy chodzi o liczbę kwiatów języczkowych w koszykach złożonych. Jako przykład podam materiał dla *Senecio fluviatilis* według kultury Ogrodu Botanicznego w Dublanach (tab. 6,4). W ust. 7 będą podane dwa podobne zliczenia dla *Chrysanthemum segetum* (tab. 7,9). U zwierząt takie zjawiska są rzadkie. Warto jest przytoczyć w tym względzie zliczenie A. G. M a y e r a odnośnie do liczby promieni u meduzy *Pseudoclytia pentata* (tab. 6,5).

Tabela 6,3

Waga dorosłych Anglików.

Waga w funtach angielskich	Ilość osób o takiej wadze
90—	2
100—	34
110—	152
120—	390
130—	867
140—	1623
150—	1559
160—	1326
170—	787
180—	476
190—	263
200—	107
210—	85
220—	41
230—	16
240—	11
250—	8
260—	1
270—	0
280—	1
Ogółem	7749

Tabela 6,4

Liczby kwiatów języczkowych w szczytowych koszykach *Senecio fluviatilis*.

Liczby kwiatów	Ich częstości
5	1
6	14
7	43
8	113
9	4
10	1
Ogółem	176

Tabela 6,5

Liczby promieni u meduzy *Pseudoclytia pentata*.

Liczby promieni	Ich częstości
2	1
3	8
4	56
5	860
6	64
7	6
8	1
Ogółem	996

Podam jeszcze kilka przykładów dwubocznych szeregów z różnych dziedzin. Bardzo ciekawe są m. i. materiały meteorologiczne a to przez swój rozmaity charakter. Dają one krzywe zmienności symetryczne, skośne, jednoboczne i wklęsłe. I tak średnie temperatury powietrza w Krakowie w lipcu dają szereg symetryczny, natomiast temperatury styczniowe — szereg bardzo silnie skośny (tab. 6,6 i 7). W obu przypadkach zliczenia opierają się na 25-letnich obserwacjach z lat 1886—1910. Temperatury są w nich zaokrąglone do całych stopni, zatem klasa k obejmuje temperatury od $k - 0.4^0$ do $k + 0.5^0$. Inne przykłady materiałów meteorologicznych będą przytoczone w dalszych częściach tego ustępu.

Tabela 6,6

Średnie temperatury dobowe lipca w Krakowie (1886—1910)

Temperatury	Ich częstości	Częstości zgrupowane w klasy po 3^0
11	1	
12	7	35
13	27	
14	38	
15	66	194
16	90	
17	97	
18	95	211
19	109	
20	90	
21	66	197
22	41	
23	21	
24	14	45
25	10	
26	2	3
27	1	
Ogółem	775	775

Tabela 6,7

Średnie dobowe temperatury stycznia w Krakowie (1886—1910)

Temperatury	Ich częstości	Częstości zgrupowane po 3°
— 28	1	
— 27	0	1
— 26	0	
— 25	0	
— 24	1	1
— 23	0	
— 22	2	
— 21	1	6
— 20	3	
— 19	6	
— 18	4	12
— 17	2	
— 16	8	
— 15	9	23
— 14	6	
— 13	13	
— 12	18	50
— 11	19	
— 10	18	
— 9	28	69
— 8	23	
— 7	43	
— 6	28	110
— 5	39	
— 4	45	
— 3	42	131
— 2	44	
— 1	63	
0	52	195
+ 1	80	
+ 2	82	
+ 3	41	153
+ 4	30	
+ 5	15	
+ 6	7	23
+ 7	1	
+ 8	1	
+ 9	0	1
+ 10	0	
	Ogółem	775

Bardzo ciekawy szereg rozdzielczy zestawił J. Witkowski dla błędów obserwacyj astronomicznych[1]). Są to różnice między kolejnymi obserwacjami rektascencji gwiazd, dokonanymi przez J. Kowalczyka w Warszawie. Materiał został zgrupowany w klasy szerokości 0.15 sekundy (tab. 6,8). Dzięki temu, że liczba wartości wariantów w klasach jest nieparzysta, środkowe wartości w tym przykładzie przypadają na realne, nie fikcyjne wartości zmiennej ewentualnej.

Tabela 6,8

Różnice między kolejnymi obserwacjami rektascencji gwiazd.

Środki klas	Ilości przypadków
—0,90	0
—0,75	55
—0,60	245
—0,45	854
—0,30	1358
—0,15	1948
—0,00	2297
+0,15	2052
+0,30	1582
+0,45	1020
+0,60	351
+0,75	70
+0,90	13
Ogółem	11854

Przytaczam dalej przykład z dziedziny zoologii. Dotyczy on liczby promieni w płetwach ogonowych flądry (*Pleuronectes*), złowionych koło Skagen i zbadanych przez C. G. Petersena (tab. 6,9). Zapożyczam go z książki Johannsena. Asymetria szeregu jest dodatnia.

[1]) J. Witkowski. An astronomical example of a non-Gaussian frequency curve. — Bull Acad. Polon. Sc. et L. Série A. 1923.

Tabela 6,9

Liczby promieni w płetwàch ogonowych flądry.

Liczby promieni	Ilości osobników
47	5
48	2
49	13
50	23
51	58
52	96
53	134
54	127
55	111
56	74
57	37
58	16
59	4
60	2
61	1
Ogółem	703

Przechodzimy do szeregów jednobocznych. Już w ust. 5 przytoczyłem przykład takiego szeregu dla liczby płatków u przylaszczki (*Anemone Hepatica*), oparty na czterdziestkowej próbie. Podaję poniżej zestawienie obfitszego materiału (tab. 6, 10). W tej tabeli podaję także podobne dane dla jaskrów (*Ranunculus*) według moich obserwacyj w okolicach Zimnej Wody i Suchowoli pod Lwowem. Zbierałem kwiaty w odległości kilku metrów jeden od drugiego, ażeby w miarę możności reprezentowały inne osobniki a nie klony. Z każdego osobnika brałem jeden kwiat. Z jaskrów brałem tylko kwiaty szczytowe, gdyż u nich zaznacza się podobna różnica między różnie położonymi kwiatami, jak u maku (por. ust. 2). Z przylaszczką nie było tego kłopotu, bo kwiaty są pojedyńcze.

Tabela 6,10

Częstość ilości płatków u jaskrowatych.

Nazwa rośliny	Ilość obserwacyj	Ilości płatków						
		5	6	7	8	9	10	11
Anemone hepatica	439	—	204	139	61	24	9	2
Ranunculus polyanthemos	428	410	16	1	—	1	—	—
R. acer	518	458	47	7	3	2	—	—
R. repens	413	314	56	26	13	3	1	—
R. flammula	145	111	24	8	2	—	—	—

Co do jaskrów warto jest zaznaczyć, że w ich kwiatach zmienność liczby owocków jest zgoła inna niż płatków: otrzymuje się dwuboczne szeregi symetryczne. Naprzykład dla *R. acer* otrzymałem w Krynicy wynik zliczeń zestawiony w tab. 6,11.

Tabela 6,11

Liczby owocków w szczytowych kwiatach *Ranunculus acer.*

Liczby owocków	Ich częstości
15—	1
20—	12
25—	36
30—	48
35—	35
40—	18
45—	2
50—	3
55—	2
Ogółem	157

Różnica w zmienności różnych cech tego samego organizmu jest tu jaskrawa, o wiele silniejsza niż w omówionym poprzednio przypadku wzrostu i wagi ludzi.

Jednoboczne krzywe zmienności nie są rzadkie w meteorologii. Naprzykład Charlier podał bardzo ciekawy przykład odnośnie do ilości dni z burzami (wyładowaniami elektrycznymi) dla Lundu za okres 105-letni (1753-1857). Liczby tabeli 6,12 podają, ile było lat, w których dany miesiąc wykazywał tyle a tyle dni z burzami.

Tabela 6,12

Ilości lat w których poszczególne miesiące miały tyle a tyle dni z burzami. Obserwacje w Lund z lat 1753—1857.

Dni z burzami	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
0	105	105	101	82	48	27	12	24	62	93	102	105
1	—	—	4	17	36	36	25	26	24	11	3	—
2	—	—	—	6	14	26	24	19	9	1	—	—
3	—	—	—	—	5	9	15	13	5	—	—	—
4	—	—	—	—	2	5	14	9	5	—	—	—
5	—	—	—	—	—	2	9	6	—	—	—	—
6	—	—	—	—	—	—	2	5	—	—	—	—
7	—	—	—	—	—	—	1	2	—	—	—	—
8	—	—	—	—	—	—	1	—	—	—	—	—
9	—	—	—	—	—	—	1	—	—	—	—	—
10	—	—	—	—	—	—	1	—	—	—	—	—
11	—	—	—	—	—	—	—	1	—	—	—	—

Podaję wreszcie sławny przykład Bortkiewicza, dotyczący częstości wypadków śmierci od kopnięcia konia w 10 korpusach armii pruskiej w poszczególnych latach 20-letniego okresu (tab. 6,13). Liczby tej tabeli podają, ile było korpusów, w których zdarzyło się tyle a tyle przypadków śmierci w ciągu roku.

Tabela 6,13

Wypadki śmierci od kopnięcia konia.

Ilości wypadków w ciągu roku	Ich częstości
0	190
1	65
2	22
3	3
4	1
Ogółem	200

W końcu trzeba omówić szeregi rozdzielcze wklęsłe. W nich najczęstsze są skrajne wartości wariantów: najmniejsze i największe. Tylko stopnie zachmurzenia dają wyraźne szeregi tego rodzaju. Na przykład dla Bazylei otrzymano następujący szereg według obserwacyj z okresu 1907—10 (tab. 6, 14).

Yule podaje jeszcze przykład ze statystyki amerykańskiej, dotyczący ilości głuchoniemych wśród dzieci rodziców, z których przynajmniej jedno było dotknięte tym kalectwem, ale sam przyznaje, że materiał jest zbyt szczupły, by zapewnić szeregowi wyraźną formę. To też pomijam ten przykład.

Tabela 6,14

Częstości stopni zachmurzenia w Bazylei dla lat 1907—1910.

Stopnie zachmurzenia	Ich częstości względne
0,0	0,138
0,1	0,062
0,2	0,051
0,3	0,047
0,4	0,039
0,5	0,034
0,6	0,033
0,7	0,048
0,8	0,052
0,9	0,078
1,0	0,418
Suma	1,000

7. Szeregi rozdzielcze złożone. Ma się je wtedy, kiedy materiał statystyczny jest mieszaniną wariantów, pochodzących z dwu albo kilku różnych populacyj. Jest to zjawisko bardzo nieprzyjemne, bo nieraz złożoność szeregu nie przejawia się w jego charakterze. Tak się dzieje, kiedy są zmieszane dwa materiały symetryczne mniej więcej jednakowo liczne o podobnym charakterze, ale z częstościami przesuniętymi w stronę mniejszych lub większych wartości wariantów. Naprzykład mamy w tabeli 7,1 zestawione takie dwa fikcyjne szeregi (1) i (2), z których drugi ma częstości przesunięte o 2 wyrazy w stronę większych wartości wariantów. Podaję tu, tak samo jak w dalszych zestawieniach, tylko ciągi częstości z pominięciem wartości wariantów, które są obojętne.

Tabela 7,1

1	8	28	56	**70**	56	28	8	1	—	—	(1)
—	—	1	8	28	56	**70**	56	28	8	1	(2)
1	8	29	64	98	**112**	98	64	29	8	1	(3)

Częstości wartości modalnych są zaznaczone tłustym drukiem. Złączone ze sobą, te dwa materiały dają szereg (3) tak samo symetryczny.

Jeżeli jeden z takich szeregów będzie bardziej liczny, powstanie szereg skośny. Tak wypadnie naprzykład, jeżeli drugi ze składanych szeregów będzie trzy razy liczniejszy (tab. 7,2).

Tabela 7,2

1	8	28	56	**70**	56	28	8	1	—	—
—	—	3	24	84	168	**210**	168	84	24	3
1	8	31	80	154	224	**238**	176	85	24	3

Tym bardziej tak będzie, jeżeli składane szeregi mają nietylko różną liczebność, ale także różne stosunki częstości, po-

zostając zresztą symetryczne. Naprzykład tak będzie w zestawieniu tabeli 7,3.

Tabela 7,3

1	8	28	56	**70**	56	28	8	1
—	—	1	6	15	**20**	15	6	1
1	8	29	62	**85**	76	43	14	2

Niewątpliwie nie jeden skośny szereg zawdzięcza tę cechę swojej złożoności. Nie można tego jednak twierdzić napewno. Natomiast jeżeli szereg ma dwie albo więcej wartości wariantów o częstościach większych niż sąsiednie, jeżeli innymi słowami ma dwa lub więcej „szczytów", to jego złożoność nie ulega wątpliwości. Musi to być zresztą szereg o dużej liczebności, bo przeskoki w częstościach, o których była mowa w ust. 4, mogą tworzyć szczyty iluzoryczne.

Tego rodzaju szereg otrzymamy naprzykład, jeżeli do szeregu (1) dodamy drugi o częstościach przesuniętych nie o 2 wyrazy, lecz o 4 (tab. 7,4).

Tabela 7,4

1	8	26	56	**70**	56	28	8	1	—	—	—	—
—	—	—	—	1	8	28	56	**70**	56	28	8	1
1	8	28	56	**71**	64	56	64	**71**	56	28	8	1

Przytoczę w dalszym ciągu realne przykłady złożonych szeregów. I tak Ludwig[1]) ogłosił następujące zestawienie liczb promieni w szczytowych baldachach podagrycznika (*Aegopodium Podagraria*), rośliny z rodziny baldaszkowych (tab 7,5). Wystąpiło tu zjawisko bardzo ciekawe: szczyty oddalone od siebie o trzy wyrazy szeregu. Autor tej książki wyka-

---

[1]) F. Ludwig. Ueber Variationskurven und Variationsflächen der Pflanzen. — Botan. Centralbl. Vol. 64 (1895).

zał, że to zjawisko jest bardzo rozpowszechnione w rodzinie baldaszkowatych [1]).

Tabela 7,5

Liczby promieni w szczytowych baldachach podagrycznika.

8	9	10	11	12	13	14	15	16	17
8	13	22	34	40	46	92	**126**	120	127
18	19	20	21	22	23	24	25	26	27
168	120	119	**136**	85	47	32	34	21	8
28	29	30	31	32					
4	2	1	1	1					

Nie mniej ciekawy jest przykład drugi, który zawdzięczamy również L u d w i g o w i [2]). Dotyczy on liczby kwiatów w kwiatostanach pierwiosnka (*Primula officinalis*). Szczytów jest aż pięć! (tab. 7,6).

Tabela 7,6

Liczby kwiatów w kwiatostanach pierwiosnka.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	13	**78**	54	**237**	173	116	**188**	76	**108**	56	19	**21**	10	6
16	17	18	19	20	21	22								
4	5	3	2	0	0	1								

Cztery z pomiędzy tych szczytów a mianowicie 3, 5, 8 i 13 są reprezentowane przez tzw. szereg F i b o n a c c i'ego nazwany tak na cześć włoskiego matematyka z XIII wieku. Sze-

[1]) D. S z y m k i e w i c z. Observations biométriques. I. — Acta Soc. Bot. Polon. Vol. 9 (1932).

[2]) F. L u d w i g. Eine fünfgipfelige Variationskurve. — Ber. Deutsch. Botan. Ges. Vol. 14 (1896).

reg ten zaczyna się od jedynki i dwójki a jego dalsze wyrazy tworzą się przez dodawanie dwóch poprzednich. Piąty szczyt ma wartość także jednej z liczb Fibonacci'ego, ale pomnożonej przez dwa: $10 = 5 \times 2$. Tego rodzaju liczby bardzo często charakteryzują wartości modalne u roślin. Już poprzednio mieliśmy 8 jako wartość modalną w okrywie sałatnicy (tab. 6,2) i dla kwiatów języczkowych *Senecio fluviatilis* (tab. 6,4). Zjawisko to nosi miano prawa Ludwiga.

Jako przykład szeregu wieloszczytowego z dziedziny zoologii przytoczę pomiary wyrostków odwłoka u skorka (*Forficula*) wykonane przez Batesona[1]) (tab. 7,7).

Tabela 7,7

Częstość długości wyrostków odwłoka u skorka, wymierzonych w mm.

3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
64	**125**	52	7	12	24	42	42	**90**

7.5	8.0	8.5	9.0
68	44	8	6

Różnice w biologicznych materiałach, powodujące wytworzenie wieloszczytowych szeregów rozdzielczych, mogą wynikać ze zróżnicowania organizmów na gatunki elementarne. I tak de Vries[2]) zbadał jednoroczną roślinę z rodziny złożonych *Chrysanthemum segetum*. Dała ona dla liczby kwiatów języczkowych w szczytowych koszykach krzywą zmienności ze szczytami przy 13 i 21 liczbach Fibonacciego (tab. 7,8).

---

[1]) Bateson. Materials for the study of variations — London. 1894.

[2]) H. de Vries. Ueber Curvenselection bei Chrysanthemum segetum. — Ber. Deutsch. Botan. Ges. Vol. 17 (1899).

Tabela 7,8

Kwiaty języczkowe u *Chrysathemum segetum.*

8	9	10	11	12	13	14	15	16	17
7	3	3	5	14	**153**	77	60	55	31

18	19	20	21	22	23	24	25	26
33	39	41	**56**	10	1	0	0	1

Przez selekcję udało się temu badaczowi wyosobnić dwa gatunki elementarne o jednoszczytowych krzywych zmienności dla liczby kwiatów języczkowych. Jeden z nich dał szereg ze szczytem przy 13, drugi przy 21 (tab. 7,9).

Tabela 7,9

Kwiaty języczkowe u *Chrysanthemum segetum*

9	10	11	12	13	14	15	16	17	18
1	3	8	31	**221**	50	8	5	4	3

19	20	21
1	2	1

14	15	16	17	18	19	20	21	22	23
1	3	0	3	7	14	43	**142**	43	21

24	25	26	27	32
11	5	3	1	1

ROZDZIAŁ II.

PODSTAWOWE CHARAKTERYSTYKI MATERIAŁÓW STATYSTYCZNYCH

8. **Wstępne uwagi.** Zgodnie z ogólnym zadaniem statystyki najważniejszym zagadnieniem badania populacyj jest wyszukanie możliwie mało licznych wielkości, któreby je charakteryzowały w sposób możliwie najbardziej poglądowy i dokładny zarazem. Są to tzw. charakterystyki. Zadaniem tego rozdziału jest omówienie podstawowych wielkości tego rodzaju, mianowicie wartości modalnej, średniej arytmetycznej i obliczonych względem niej momentów.

Z tych charakterystyk podstawowych wypadnie wyprowadzić jeszcze dalsze, pochodne jak to się mówi: odchylenie przeciętne i średnie, skośność i inne, co będzie przedmiotem następnego rozdziału.

9. **Wartość modalna.** Jest to najczęstsza wartość wariantów, o której już była mowa w poprzednim rozdziale. Najłatwiej daje się ona ustalić dla szeregów rozdzielczych o niegrupowanych wartościach wariantów. I tak w ust. 6 mieliśmy liczne przykłady: dla liczby kwiatów w koszykach sałatnicy wartość modalna była 17, dla liczby wewnętrznych listków okrywy u tejże rośliny 8 (tab. 6,2), dla liczby promieni u meduzy *Pseudoclytia pentata* 5 (tab. 65), dla liczby promieni w płetwach ogonowych flądry 53 (tab. 6,9) itd. U roślin ta wartość bardzo często równa się liczbom szeregu Fibonacci'ego, o czym już była mowa w ust. 6.

Ustalenie wartości modalnej dla szeregów o zgrupowanych wariantach jest o wiele trudniejsze. Łatwo jest wprawdzie wskazać klasę o największej częstości, ale to nie daje żadnej określonej liczby, co najwyżej określa granice, w których się ona mieści. Chcąc w tym przypadku wyznaczyć określoną wartość, trzeba wyznaczyć częstości wartości wariantów za pomocą ciągłej funkcji i wybrać największą wartość tej funkcji z pomiędzy jej wartości dla poszczególnych wartości wariantów. Odpowiadająca jej wartość zmiennej ewentualnej będzie wartością modalną. Zadanie to będzie rozpatrzone w rozdziale VI. Tam w ust. 36 tab. 4 będzie tytułem przykładu wyznaczona wartość modalna dla liczby kwiatów w szczytowych koszykach bławatka. Wypada ona równa 34, liczbie z szeregu Fibonacci'ego.

Wyznaczenie wartości modalnej wymaga dużego materiału, tym większego im większa jest ta wartość. Wobec tego na przykład wartość modalna 34 dla bławatka przy 300 wariantach jest niepewna, tak samo wartość modalna 53 dla flądry przy 703 wariantach, natomiast wartość 8 dla kwiatów języczkowych w koszykach *Senecio fluviatilis* jest ustalona pomimo 176 zaledwie wariantów (tab. 6,4).

Jeżeli szereg jest wieloszczytowy, szczytowe wartości wariantów mogą być uważane za pewnego rodzaju wartości modalne.

10. Średnia arytmetyczna i geometryczna. Średnia arytmetyczna, od której zaczniemy, jest w zespole charakterystyk wielkością podstawową, bo prawie wszystkie charakterystyki są z nią w ten czy inny sposób związane. Jak to powszechnie jest wiadome, średnia arytmetyczna jest ilorazem od podzielenia sumy wariantów przez ich liczbę. Jeżeli zatem oznaczymy warianty przez symbole $u_1, u_2, u_3, \ldots u_n$, gdzie n jest ich liczba, to możemy napisać że ich średnia arytmetyczna m równa się

$$m = \bar{u} = \frac{u_1 + u_2 + \ldots + u_n}{n}$$

Oznaczyliśmy tu średnią arytmetyczną symbolem $\bar{u}$, to znaczy symbolem wariantów z kreską u góry. Takie znakowanie jest bardzo wygodne, bo wskazuje odrazu z jakich wielkości średnia arytmetyczna została wyprowadzona.

Będziemy w tym zagadnieniu, jak i w wielu dalszych, posługiwali się symbolem Σ. Oznacza on sumę wielkości, stanowiących pewien zespół albo sumę wartości jakiejś funkcji takich wielkości. Wielkości te oznacza się pewną literą z dodaniem u dołu litery i, zaznaczającej numerację rozpatrywanych wielkości. Po znaku Σ pisze się symbol danych wielkości albo symbol funkcji. W ten sposób sumowanie wariantów dla obliczenia średniej arytmetycznej może być oznaczone przez wyrażenie Σu_i. Dla bliższego określenia, które wielkości są sumowane, pisze się u dołu litery Σ numer pierwszej z nich, u góry zaś numer ostatniej. Naprzykład gdyby chodziło o zsumowanie wariantów od drugiego do ósmego, napisalibyśmy $\sum_{2} u$. Dla obliczenia średniej arytmetycznej sumuje się wszystkie n wariantów. Wobec tego wzór dla tej charakterystyki będzie

$$m = \bar{u} = \frac{\sum_{1}^{n} u_i}{n} \qquad (1)$$

W związku z użyciem symbolu Σ należy pamiętać o tym, że sumowanie może być algebraiczne i arytmetyczne. Pierwsze wykonuje się w ten sposób, że dodaje się osobno dodatnie i ujemne składniki i od sumy dodatnich odejmuje się sumę ujemnych. Wynik może wypaść dodatni lub ujemny. Przy sumowaniu algebraicznym symbole sumowanych wielkości wpisuje się za znakiem Σ bez żadnych szczególnych odznak. Inaczej jest przy sumowaniu arytmetycznym — tu dodaje się razem wszystkie liczby bez względu na znak, uważając je wszystkie za dodatnie. Innymi słowami sumuje się absolutne wartości liczb. W tym przypadku symbol sumowanych wielkości za znakiem Σ ujmuje się w nawias

złożony z pionowych kresek, np. w ust. 17 spotkamy się z wyrażeniem $\sum_{1}^{n} \left| u_i - \overline{u} \right|$. Taki nawias oznacza absolutną wartość danej wielkości. Jeżeli nie będzie to specjalnie zaznaczone, wszystkie sumy w tej książce należy rozumieć jako algebraiczne.

Wartość średniej arytmetycznej jest pośrednia między wartościami wariantów. Stanowi ona pewnego rodzaju ośrodek w ich zbiorze. Nazwa jej podkreśla tę właściwość. Dla skrócenia nazywa się ją nieraz po prostu średnią.

Jeżeli materiał statystyczny jest niewielki, obliczenie średniej arytmetycznej nie sprawia żadnych trudności. Może tylko chodzić o dokładność wyników. Otóż można przyjąć za zasadę przy wszelkich obliczeniach, że wynik nie powinien zawierać więcej niż cztery cyfry. Zwykle wystarczają nawet trzy. W tym prawidle zera przed znaczącymi cyframi nie liczą się, liczą się natomiast jeżeli występują na końcu. Naprzykład liczba 0.02750 ma cztery cyfry, nie sześć, zresztą także nie trzy. Powyższe prawidło dotyczy tylko ostatecznego wyniku. Natomiast obliczenia trzeba nieraz przeprowadzić z większą ilością znaków dziesiętnych. Na ogół trzeba mieć o dwa więcej niż ich ma być w wyniku ostatecznym, by otrzymać dokładną cyfrę dla ostatniego znaku.

Warto jest jeszcze przy rachunkach zwrócić uwagę na następujący szczegół. Wypada często w wynikach odrzucać dalsze znaki dziesiętne. Przyjęto powszechnie, że jeżeli odrzucone cyfry dają liczbę mniejszą od połowy jednostki ostatniej zachowanej cyfry albo równą połowie, to tę cyfrę zostawia się bez zmiany, w przeciwnym zaś razie powiększa się ją o jedynkę. Ustalony w ten sposób wynik jest zatem nieco mniejszy albo nieco większy od dokładnej wartości. Otóż dobrze jest wiedzieć, czy jest on mniejszy od niej, czy większy. W tym celu należy ostatnią cyfrę wyniku podkreślać, jeżeli jest on większy. I tak, naprzykład, jeżeli zamiast 5.286 weźmiemy 5.29, to należy 9 podkreślić.

Użyteczność takiego znakowania ujawnia się nieraz, jeżeli wynik podlega dalszym rachunkom. Przypuśćmy, że podaną powyżej liczbę 5.29 trzeba podzielić przez dwa z zachowaniem tej samej liczby znaków dziesiętnych. Będziemy wtedy w kłopocie, czy mamy wziąć 2.64 czy 2.65, jeżeli nie będziemy wiedzieli, czy ta liczba 5.29 była obliczona z nadmiarem, czy z niedostatkiem. W pierwszym przypadku iloraz będzie 2.64, w drugim 2.65.

Użyteczność omawianego znakowania ujawnia się także wtedy, kiedy trzeba odrzucić ostatnią cyfrę a nią jest 5, naprzykład gdy chodzi o liczbę 7.385. Czy wziąć 7.38 czy 7.39? Oczywiście jeżeli to będzie 7.385, trzeba będzie wziąć 7.38, jeżeli zaś 7.385 bez podkreślenia — to 7.39. Takie piątki na końcu liczb są bardzo nieprzyjemne i zmuszają nieraz do obliczania dalszych znaków dziesiętnych.

Jeżeli materiał statystyczny jest duży, obliczanie średniej arytmetycznej jest bardziej zawiłe. Materiał ma wtedy formę szeregu rozdzielczego. Przypuśćmy, że wartości wariantów są w liczbie $l : x_1, x_2, x_3, .., x_l$. Stanowią one zmienną ewentualną, o której była mowa w ust. 5. Niech ich częstości będą odpowiednio $f_1, f_2, f_3, \ldots\ldots, f_l$, przy czym suma tych częstości równa się n, liczbie wariantów. Wtedy średnia arytmetyczna zostanie określona równaniem:

$$m = \bar{u} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_l x_l}{f_1 + f_2 + \ldots + f_l} = \frac{\sum_1^l f_i x_i}{\sum_1^l f_i} = \frac{\sum_1^l f_i x_i}{n} \qquad (2)$$

Średnią arytmetyczną wariantów można uważać za średnią arytmetyczną wartości wariantów, ale z tym zastrzeżeniem, że to będzie średnia w a ż o n a. Tak nazywa się średnia arytmetyczna liczb, różniących się między sobą pod względem swojego znaczenia w danym zagadnieniu. Wtedy każda taka liczba ma swoją „wagę", określającą to znaczenie, a średnią arytmetyczną otrzymuje się, mnożąc każdą liczbę przez jej wagę i dzieląc sumę takich iloczynów przez sumę wag. Tak właśnie przedstawia się wzór (2) podany powyżej. W nim rolę wag dla

wartości zmiennej ewentualnej x odgrywają częstości f tych wartości. Średnią ważoną oznacza się podobnie jak zwykłą średnią przez symbol składników z kreską u góry, ale podwójną. Można więc napisać:

$$m = u = \bar{\bar{x}}$$

Ze średnią arytmetyczną ważoną ma się do czynienia m. i. wtedy, kiedy trzeba przeprowadzić obliczenia, nie mając bezpośrednio wariantów, mając tylko częściowe średnie arytmetyczne, to znaczy jeżeli zbiór wariantów został podzielony na części i dla każdej części z osobna była obliczona średnia. Wagą każdej takiej częściowej średniej będzie liczba wariantów, z których ona była obliczona. I tak przypuśćmy, że mamy do dyspozycji trzy średnie częściowe: jedną o wartości 5.2 dla 12 wariantów, drugą 7.1 dla 20 wariantów i trzecią 9.0 dla 5 wariantów. Średnią arytmetyczną ważoną obliczymy według wzoru

$$\frac{5.2 \times 12 + 7.1 \times 20 + 9.0 \times 5}{12 + 20 + 5} = 6.7 .$$

Byłoby natomiast błędnym obliczanie średniej arytmetycznej w tym przypadku według wzoru

$$\frac{5.2 + 7.1 + 9.0}{3} = 7.1,$$

gdyż omawiane wielkości, dla których ma być znaleziona średnia, nie są równoważne: największe znaczenie ma 7.1 reprezentująca 20 wariantów, najmniejsze 9.0, reprezentująca tylko 5.

Równania (1) i (2) mogą służyć do obliczenia średniej arytmetycznej. Lepiej jest jednak przy dużej ilości wariantów uprościć sobie zadanie przez zastosowanie *metody wartości wyjściowej*. Metoda ta opiera się na podstawowej własności średniej arytmetycznej, którą można sformułować w sposób następujący: *suma algebraiczna odchyleń*

wariantów od ich średniej równa się zeru. Odchyleniami nazywamy różnice między wariantami a jakąś stałą wielkością. Jeżeli tę wielkość oznaczymy przez a, wzór na odchylenie będzie $u_i - a$.

Dla przeprowadzenia dowodu weźmiemy zasadnicze równanie

$$\bar{u} = \frac{\sum_1^n u_i}{n}$$

Przekształcając je kolejno otrzymujemy:

$$n\bar{u} = \sum_1^n u_i, \quad \sum_1^n \bar{u} = \sum_1^n u_i, \quad \sum_1^n (u_i - \bar{u}) = 0.$$

Ostatnia forma równania jest matematycznym wyrazem omawianej własności średniej arytmetycznej.

Tylko średnia arytmetyczna ma taką własność. Suma algebraiczna odchyleń od każdej innej liczby nie jest równa zeru. W istocie, niech będzie a jakakolwiek liczba. Będziemy mieli

$$\sum_1^n (u_i - a) = \sum_1^n (u_i - \bar{u} + \bar{u} - a) = \sum_1^n (u_i - \bar{u}) + \sum_i^n (\bar{u} - a),$$

Pierwsza z otrzymanych dwóch sum jest sumą odchyleń od średniej i przeto równa zeru. Druga zaś równa się $n\ (\bar{u} - a)$. Wobec tego otrzymamy równanie

$$\sum_1^n (u_i - a) = n(\bar{u} - a),$$

które jest dowodem rozpatrywanego twierdzenia.

Z powyższego równania można dla średniej arytmetycznej otrzymać wzór:

$$\bar{u} = a + \frac{\sum_1^n (u_i - a)}{n} \tag{3}$$

który jest podstawą metody wartości wyjściowej dla obliczania tej charakterystyki. W nim a jest wartością wyjściową.

Tę wielkość można obrać dowolnie. Dla uproszczenia bierze się liczbę całkowitą pośrednią między wartościami wariantów. Wtedy w obliczeniach będziemy mieli do czynienia z liczbami stosunkowo małymi.

Metody wartości wyjściowej można używać przy każdej formie materiału statystycznego. Najważniejsze jest jednak jej zastosowanie do szeregów rozdzielczych. *Wtedy wybiera się jako wartość wyjściową najczęstszą wartość wariantów, przez co liczne odchylenia staną się równe zeru.*

Głównym zadaniem jest teraz obliczenie sumy odchyleń wariantów od wartości wyjściowej czyli sumy $\sum_{1}^{n}(u_i - a)$ w równaniu (3). Będzie ona równa sumie

$$f_1(x_1 - a) + f_2(x_2 - a) + \;.\;.\;.\; f_l(x_l - a) = \sum_{1}^{l} f_i(x_i - a)$$

i równanie (3) przybierze formę:

$$\bar{u} = a + = \frac{\sum_{1}^{l} f_i(x_i - a)}{\sum_{1}^{i} f_i}$$

W nim ostatni wyraz przyjęto oznaczać przez *b*:

$$b = \frac{\sum_{1}^{l} f_i(x_i - a)}{\sum_{1}^{l} f_i} = \frac{\sum_{1}^{l} f_i(x_i - a)}{n}$$

Dodawanie wielkości *b* do wartości wyjściowej daje średnią arytmetyczną : $u = a + b$.

Obliczanie średniej arytmetrycznej dla szeregu rozdzielczego należy przeprowadzać według schematu, który pokażę na przykładzie liczby kwiatów w koszykach sałatnicy (tab. 6,2).

Obieramy jako wartość wyjściową wartość modalną 17. Otrzymamy wtedy (tab. 10,1)

Tabela 10,1

Liczby kwiatów w koszykach sałatnicy

x	f	$x - a$	$f(x - a)$
11	2	—6	—12
12	7	—5	—35
13	16	—4	—64
14	29	—3	—87
15	42	—2	—84
16	45	—1	—45
			—327
17	54	0	
18	46	+2	+46
19	43	+2	+86
20	36	+3	+108
21	19	+4	+76
22	15	+5	+75
23	1	+6	+6
24	2	+7	+14
25	1	+8	+16
	359		+427

Wielkość b wypadnie równa $\frac{427 - 327}{359} = 0.279$ a średnia arytmetyczna $m = u = a + b = 17 + 0.279 = 17.279$. Wystarczy wziąć cztery cyfry i przyjąć $m = 17.28$.

W podobny sposób przeprowadza się obliczenia dla szeregów rozdzielczych z wariantami zgrupowanymi w klasy.

T r a k t u j e s i ę w t e d y m a t e r i a ł s t a t y s t y c z n y, j a k g d y b y w a r i a n t y k a ż d e j k l a s y m i a ł y j e d n a k o w ą w a r t o ś ć, r ó w n ą n u m e r o w i d a n e j k l a s y. Numery te, będące nową formą zmiennej ewentualnej, oznacza się dla odróżnienia od wartości pierwotnej zmiennej ewentualnej symbolem x' zamiast x i operuje się nimi zupełnie tak samo jak poprzednio z niezmienionymi wartościami wariantów. Tak samo częstości klas oznacza się przez f'.

Wyniki otrzymuje się wyrażone w szerokości klas jako jednostce pomiarowej. Nie są one ostateczne. Dla oznaczenia ich używa się tych samych symbolów co poprzednio, ale z dodaniem u góry przecinka, tak samo, jak dla oznaczenia nowej formy zmiennej ewentualnej, a więc przez b' zamiast b, m' zamiast m. Ostateczne wyniki otrzymuje się, mnożąc tymczasowe przez szerokość klas.

Tytułem przykładu obliczymy średnią arytmetyczną liczby kwiatów w szczytowych koszykach bławatka z Zimnej Wody. Weźmiemy do tego zadania z ust. 5 wszystkie trzy formy (4), (5) i (6) materiału zgrupowanego po trzy wartości wariantów. Zrobimy to dlatego, żeby wykazać wpływ różnego grupowania na wynik obliczeń. Wyniki bowiem są obciążone błędami, gdyż warianty zawarte w tej samej klasie mają częściowo różne wartości. Błędy te są nieuniknione i nieraz bardzo przykre. Chcąc ich uniknąć, trzeba oprzeć się na pierwotnym materiale, co pociąga za sobą wielkie utrudnienie w rachunkach. Korzyści, które się osiągnie, są zwykle zbyt małe, by warto było sobie ten trud zadawać. Trzeba jednak zawsze zdawać sobie sprawę z wielkości błędów powodowanych przez grupowanie wariantów. Właśnie dla zobrazowania tego zagadnienia przerobimy wszystkie trzy formy naszego szeregu i nadto porównamy otrzymane przybliżone wyniki z dokładnymi, wyprowadzonymi z pierwotnej (2) formy szeregu.

Dla pierwszej formy szeregu obieramy jako wartość wyjściową x, dla najliczniejszej (modalnej) klasy 31—33, a więc $a = 10^2/_3$. Otrzymamy wyniki zestawione w tabeli 10,2.

Tabela 10,2

Liczby kwiatów w koszykach bławatka. Pierwszy sposób grupowania.

Zakresy klas	Wartości środkowe	x'	f'	$x' - a$	$f'(x' - a)$
19—21	20	6 ⅔	1	—4	—4
22—24	23	7 ⅔	6	—3	—18
25—27	26	8 ⅔	16	—2	—32
28—30	29	9 ⅔	43	—1	—43
31—33	32	10 ⅔	71	0	—97
34—36	35	11 ⅔	56	+1	56
37—39	38	12 ⅔	50	+2	100
40—42	41	13 ⅔	34	+3	102
43—45	44	14 ⅔	9	+4	36
46—48	47	15 ⅔	5	+5	25
49—51	50	16 ⅔	7	+6	42
52—54	53	17 ⅔	1	+7	7
55—57	56	18 ⅔	0	+8	0
58—60	59	19 ⅔	0	+9	0
61—63	62	20 ⅔	1	+10	10
			300		+378

Otrzymujemy $b' = +\frac{281}{300} = +0{,}937$ i $m' = 10\,{}^2/_3 + 0{,}937 =$ $= 10.667 + 0.937 = 11.604$ w przedziałach klasowych. Ostatecznie wypada: $m = 3 \times 11.604 = 34.812$. Można przyjąć $m = 34.81$.

Dla drugiej formy szeregu, biorąc $a = 11$, otrzymamy następujące wyniki (tab. 10,3): $b' = +\frac{187}{300} = +0{,}623$, $m' = 11 +$ $+ 0{,}623 = 11{,}623$, zatem $m = 34{,}869$. Można wziąć $m = 34{,}87$.

Wreszcie trzecia forma szeregu da nam przy $a = 11\,{}^1/_3$ (tab. 10,4): $b' = +\frac{73}{300} = +0.243$, $m' = 11\,{}^1/_3 + 0.243 =$ $= 11.333 + 0.243 = 11.576$. Stąd $m = 34.728$. Można przyjąć $m = 34.73$.

Obliczając dokładną wartość średniej arytmetycznej na podstawie szeregu (2) ust. 5, otrzymuje się liczbę 34,77. Zestawiając ją z przybliżonymi wynikami, znajdziemy błędy: dla pierwszej formy szeregu + 0.40, dla drugiej + 0.10, dla trzeciej — 0.04. Najlepiej wypadły wyniki dla pierwszej i trzeciej formy szeregu. Pochodzi to stąd, że w pierwszej formie szeregu częstości zgrupowały się najbardziej w środkach klas, a trzecia forma jest najlepiej wyrównana. Różnice są zresztą wszystkie niewielkie. Tak jest zawsze dla szeregów mniej więcej symetrycznych, takich jak rozpatrywany szereg bławatka, bo wtedy błędy pochodzące od mniejszych wariantów równoważą się mniej więcej przez błędy od większych. Dla szeregów bardzo skośnych i jednobocznych błędy mogą wypaść duże.

Tabela 10,3

Liczby kwiatów w koszykach bławatka.
Drugi sposób grupowania.

Zakresy klas	Wartości środkowe	x'	f'	$x'-a$	$f'(x'-a)$
20—22	21	7	2	—4	—8
23—25	24	8	9	—3	—27
26—28	27	9	31	—2	—62
39—31	30	10	43	—1	—43
32—34	33	11	70	0	—140
35—37	36	12	53	+1	53
38—40	39	13	41	+2	82
41—43	42	14	33	+3	99
44—46	45	15	8	+4	32
47—49	48	16	4	+5	20
50—52	51	17	4	+6	24
53—55	54	18	1	+7	7
56—58	57	19	0	+8	0
59—61	60	20	0	+9	0
62—64	63	21	1	+10	10
			300		+327

Tabela 10,4

Liczby kwiatów w koszykach bławatka.
Trzeci sposób grupowania.

Zakresy klas	Wartości środkowe	x'	f'	$x'-a$	$f(x'-a)$
21—23	22	7⅓	5	—4	—20
24—26	25	8⅓	11	—3	—33
27—29	28	9⅓	38	—2	—76
30—32	31	10⅓	61	—1	—61
33—33	34	11⅓	63	0	—190
36—18	37	12⅓	49	+1	49
39—41	40	13⅓	37	+2	74
42—44	43	14⅓	22	+3	66
45—47	46	15⅓	5	+4	20
48—50	49	16⅓	4	+5	20
51—53	52	17⅓	3	+6	18
54—56	55	18⅓	1	+7	7
57—59	58	19⅓	0	+8	0
60—62	61	20⅓	1	+9	9
			300		+263

Dla lepszego wyjaśnienia metody obliczeń średniej arytmetycznej przytaczam jeszcze jeden przykład szeregu zgrupowanego, mianowicie szereg otrzymany z pomiarów wagi dorosłych Anglików (tabl. 6,3). Znaczenie tego przykładu polega na dokładnym wyjaśnieniu pojęcia wartości środkowej. Szerokość klas jest tu, tak jak to się często robi przy pomiarach, równa 10 i początkowe wartości wariantów w klasach są wielokrotne 10. Zdawałoby się, że wartości środkowe powinny być wielokrotne 5. Otóż tak nie jest. Łatwo jest przekonać się, że będą one ułamkowe, fikcyjne — zakończone cyframi 4.5, naprzykład dla klasy 90—99 wartość środkowa nie będzie 95, lecz 94.5.

Tabela 10,5

Waga dorosłych Anglików.

Podziały klasowe	Wartości środkowe	x'	f'	$x' - a$	$f'(x' - a)$
90—99	94,5	9,45	2	—5	—10
100—109	104,5	10,45	34	—4	—136
110—119	114,5	11,45	152	—3	—456
120—129	124,5	12,45	390	—2	—780
130—139	134,5	13,45	867	—1	—867
140—149	144,5	14,45	1623	0	—2249
150—159	154,5	15,45	1559	+1	1559
160—169	164,5	16,45	1326	+2	2652
170—179	174,5	17,45	787	+3	2361
180—189	184,5	18,45	476	+4	1904
190—199	194,5	19,45	263	+5	1315
200—209	204,5	20,45	107	+6	642
210—219	214,5	21,45	85	+7	595
220—229	224,5	22,45	41	+8	328
230—239	234,5	23,45	16	+9	144
240—249	244,5	24,45	11	+10	110
250—259	254,5	25,45	8	+11	88
260—269	264,5	26,45	1	+12	12
270—279	274,5	27,45	0	+13	0
280—289	284,5	28,45	1	+14	14
			7749		+11724

Wypada zatem (tabl. 10.5) $b' = + \frac{9475}{7749} = + 1.223$, $m' = 14.5 + 1.223 = 14.673$, skąd $m = 146.7$.

Jeszcze jedną rzecz trzeba podać o średniej arytmetycznej. Poznaliśmy już jej podstawową własność, że suma algebraiczna odchyleń wariantów od niej jest równa zeru. Otóż posiada ona jeszcze jedną ważną własność: suma kwadratów odchyleń od średniej arytmetycznej jest mniejsza od każdej innej liczby. Istotnie niech będzie

a dowolna liczba. Suma kwadratów odchyleń od niej będzie $\sum_1^n (u_i - a)^2$. Wstawmy do tego wzoru $a = \bar{u} - b$, gdzie b jest znaną już nam wielkością. Będziemy mieli:

$$\sum_1^n (u_i - a)^2 = \sum_1^n (u_i - \bar{u}_i + b)^2 = \sum_1^n (a\, u_i - \bar{u})^2 + \sum_1^n 2b(u_i - \bar{u}) + .$$

$$+ \sum_1^n b^2 = \sum_1^n (u_i - \bar{u})^2 + 2b \sum_1^n (u_i - \bar{u}) + nb^2$$

W otrzymanym trójmianie drugi wyraz zawiera jako czynnik sumę odchyleń od średniej, będzie zatem równy zeru. Ostatecznie wypadnie równanie

$$\sum_1^n (u_i - a)^2 = \sum_1^n (u_i - \bar{u})^2 + nb^2,$$

które stanowi dowód omawianej własności średniej arytmetycznej.

Przechodzimy wreszcie do średniej geometrycznej. Średnia geometryczna danych wielkości jest pierwiastkiem z iloczynu tych wielkości. Stopień tego pierwiastka równa się ich liczbie. W ten sposób średnia geometryczna wielkości a i b jest $\sqrt[2]{ab}$, średnia geometryczna wielkości a, b i c jest $\sqrt[3]{abc}$ itd.

Średnia geometryczna ma mniejsze znaczenie od średniej arytmetycznej, ale czasem jest niezbędna, mianowicie wtedy, kiedy jakiś stan rzeczy trzeba scharakteryzować za pomocą jednej liczby na podstawie dwóch albo kilku różnych cech. Naprzykład gdybyśmy chcieli scharakteryzować wartość kłębów ziemniaczanych na podstawie ich wielkości i zawartości skrobii, to średnia arytmetyczna nie mogłaby być użytą, gdyż dodawać można tylko wielkości tego samego rodzaju. Natomiast nic nie stoi na przeszkodzie wziąć średnią geometryczną z wagi i zawartość w nich skrobii. Podobnie „dorodność ludzi możnaby scharakteryzować przez średnią geometryczną z ich wagi i wzrostu. Z tego rodzaju średniej zrobimy użytek w ust. 40 przy omawianiu charakterystyk korelacji.

Dodatkowo trzeba wspomnieć o tym, że czasem — bardzo rzadko — używany jest jeszcze jeden rodzaj średnich — średnia harmoniczna. Nie ma ona poważniejszego znaczenia i omawiać jej nie będę.

11. Pojęcie momentów. Średnia arytmetyczna, mimo podstawowego jej znaczenia, nie wystarcza do scharakteryzowania materiału statystycznego. Stanowi ona pewnego rodzaju ośrodek w zbiorze wariantów, nie określa jednak, jak są one w nim rozmieszczone. Do tego są potrzebne odchylenia wariantów od niej, z którymi już mieliśmy do czynienia w poprzednim ustępie. Ponieważ odchylenia są różne, trzeba ich zbiór scharakteryzować za pomocą odpowiednich wielkości. Jest rzeczą zupełnie naturalną wziąć w tym celu ich średnią arytmetyczną. Jest to jednak niewystarczające.

Spotkamy się przede wszystkim z trudnością wynikającą z natury odchyleń w średniej arytmetycznej: ich suma algebraiczna jest zawsze równa zeru i średnia przeto także równa się zeru. Można temu coprawda zaradzić, traktując osobno odchylenia dodatnie i ujemne albo biorąc ich wartości absolutne. Główna rzecz jest jednak w czym innym: różne zbiory liczb mogą dać tę samą średią arytmetyczną. Przeto średnia arytmetyczna odchyleń od średniej, gdyby się nawet nie równała zeru, nie wystarcza dla charakterystyki ich zbioru, bo może się znaleźć inny materiał, który da dla zbioru odchyleń od średniej taką samą średnią. Trzeba wobec tego wyszukać inne jeszcze charakterystyki tego zbioru, któreby były różne dla różnych materiałów.

Dla rozwiązania takiego zagadnienia niema innego środka, jak użycie obok odchyleń jeszcze ich potęg: kwadratów, sześcianów itd. Naturalnie dla otrzymanych tą drogą zbiorów trzeba będzie znowu obliczać średnie arytmetyczne. W ten sposób dochodzimy do pojęcia momentów.

Momentem k-ego stopnia względem jakiejś wielkości a nazywamy średnią arytmetyczną k-ych potęg odchyleń wariantów od tej wiel-

kości. Z tego określenia wyprowadzamy ogólny wzór dla momentów

$$\mu_k = \frac{\sum_{1}^{n} (u_i - a)^k}{n} \qquad (1)$$

Z tego rodzaju wielkościami już mieliśmy do czynienia przy obliczaniu średniej arytmetycznej. I tak wielkość b, która tam występowała, jest niczym innym jak momentem pierwszego stopnia względem wartości wyjściowej i można napisać:

$$b = \frac{\sum_{1}^{n} (u_i - a)}{n} = \mu_1$$

Sama średnia arytmetyczna jest momentem pierwszego stopnia a mianowicie względem zera:

$$m = \frac{\sum_{1}^{n} (u_i - a)}{n}$$

Zgodnie z poprzednimi wywodami dla charakterystyki materiałów statystycznych weźmiemy momenty względem średniej arytmetycznej. Będziemy je nazywali po prostu momentami i będziemy oznaczali literą m z odpowiednią liczbą wskaźnikową:

$$m_k = \frac{\sum_{1}^{n} (u_i - u)^k}{n} \qquad (2)$$

Przy obliczaniu ich wypadnie posługiwać się momentami względem wartości wyjściowej. Będziemy je nazywali momentami pomocniczymi i oznaczali literą μ. Wielkość b jest pierwszym takim momentem.

Ile trzeba wziąć momentów dla należytej charakterystyki materiału statystycznego? Widzieliśmy, że moment pierwszego

stopnia nie wystarcza, nawet jeżeli go weźmiemy osobno dla dodatnich i ujemnych odchyleń. Trzeba do niego dodać moment drugiego stopnia. Charakterystyka materiału będzie lepsza, bo jeżeli jest dużo różnych materiałów o tym samym momencie pierwszego stopnia, to jest mniej takich, co mają te same momenty i pierwszego i drugiego sopnia. Jeszcze mniej będzie takich, co mają te same momenty pierwszego, drugiego i trzeciego stopnia itd.

Jak daleko należy iść tą drogą? Oczywiście nie za daleko, bo wtedy dla charakterystyki zbioru liczb, jakim jest materiał statystyczny, mielibyśmy nowy zbiór innych liczb, który też wymagałby charakterystyki. Niektórzy badacze, szczególnie meteorologowie, zadowalają się momentem pierwszego stopnia, odpowiednio zmienionym, tak żeby nie równał się zawsze zeru. Większość zatrzymuje się na momencie drugiego stopnia. Mniej liczni idą dalej. Poza czwarty stopień iść nie warto, bo jak to zobaczymy w dalszym ciągu, za pomocą pierwszych czterech momentów można dostatecznie dokładnie scharakteryzować każdy materiał statystyczny.

12. Momenty pierwszego stopnia. Są one, jak to widzieliśmy w ust. 10 i 11, równe zeru dla każdego materiału. Nic jednak nie stoi na przeszkodzie, by obliczyć taki moment osobno dla wariantów mniejszych od średniej i osobno dla wariantów większych do niej. W a r i a n t y m n i e j s z e o d ś r e d n i e j noszą nazwę m i n u s w a r i a n t ó w, w i ę k s z e o d n i e j — n a z w ę p l u s w a r i a n t ó w. Te nazwy często używane i wygodne mają niestety tę wadę, że używa się ich także w innym znaczeniu: m i n u s w a r i a n t a m i n a z y w a s i ę n i e r a z w a r i a n t y m n i e j s z e o d w a r t o ś c i m o d a l n e j, p l u s w a r i a n t a m i — w a r i a n t y w i ę k s z e o d n i e j.

Z powyższego wynika, że wypada rozpatrywać dwa c z ę ś c i o w e m o m e n t y p i e r w s z e g o s t o p n i a. Pierwszy z nich obliczony dla minus wariantów jest ujemny, drugi zaś obliczony dla plus wariantów jest dodatni. Dla obliczenia

trzeba najpierw podzielić szereg rozdzielczy na dwie części. Przypuśćmy, że pierwsza część — złożona z minus wariantów — obejmuje wyrazy szeregu od 1 do k z n_1 wariantami, druga zaś — złożona z plus wariantów — wyrazy od $k+1$ do n z n_2 wariantami.

Dla pierwszego częściowego momentu pierwszego stopnia, który oznaczamy symbolem m_{11}, będziemy mieli wzór:

$$m_{11} = \frac{f_1(x_1 - \bar{u}) + f_2(x_2 - \bar{u}) + \ldots f_k(x_k - \bar{u})}{f_1 + f_2 + \ldots + f_k}$$

Wstawiamy do niego $\bar{u} = a + b$, gdzie a i b są znanymi już nam wielkościami. Wstawiamy nadto: $f_1 + f_2 + \ldots + f = n_1$. Otrzymamy

$$m_{11} = \frac{f_1(x_1 - a - b) + f_2(x_2 - a - b) + \ldots + f_k(x_k - a - b)}{n_1} =$$

$$= \frac{\sum_1^k (x_i - a) - b(f_1 - f_2 + \ldots + f_k)}{n_1} =$$

$$= \frac{\sum_1^k (x_i - a) - n_1 b}{n_1} = -\frac{\sum_1^k (a - x_i) + n_1 b}{n_1} \quad (1)$$

Dla drugiego częściowego momentu pierwszego stopnia, który oznaczymy symbolem m_{12} w podobny sposób wyprowadzimy wzór:

$$m_{12} = \frac{\sum_{k+1} f_i(x_i - a) - n_2 b}{n_2}. \quad (2)$$

Za pomocą powyższych wzorów można przeprowadzać obliczenia tych momentów metodą wartości wyjściowej. Przy tym trzeba mieć na uwadze, że licznik w równaniu (1) jest absolutną wartością sumy ujemnych odchyleń od średniej, a licznik w równaniu (2) jest sumą dodatnich odchyleń. Ponieważ suma algebraiczna wszystkich takich odchyleń jest równa

zeru, oba liczniki powinny być sobie równe. Ewentualne różnice pochodzą z odrzucania dalszych znaków dziesiętnych.

Zastosujemy te wzory do liczby kwiatów w koszykach sałatnicy. Będziemy tu mogli skorzystać z rachunków ust. 10. Minus warianty będą miały wartości od 11 do 17, plus warianty — od 18 do 25. Pierwszych jest 195, drugich 164. Weźmiemy tę samą wartość wyjściową $a = 17$. Otrzymamy:

$$m_{11} = -\frac{327 + 195 \times 0{,}279}{195} = -\frac{327 + 54{,}4}{195} = -\frac{381{,}4}{195} =$$

$$= -1{,}96$$

$$m_{12} = +\frac{427 - 164 \times 0{,}279}{164} = +\frac{427 - 45{,}8}{164} = +\frac{381{,}2}{164} = +2{,}32$$

W podobny sposób przeprowadza się obliczenia dla materiałów zgrupowanych w klasy. Dla trzech form szeregu rozdzielczego liczby kwiatów w koszykach bławatka będziemy mieli wartości (tab. 12,1). W tej tabeli są podane wartości momentów w przedziałach klasowych, oznaczone symbolami m'_{11} i m'_{12}, oraz ostateczne ich wartości otrzymane przez pomnożenie przez szerokość klas, to znaczy przez trzy. Dokładne wartości są $m_{11} = -4{,}473$ i $m_{12} = +4.782$.

Tabela 12,1

Momenty pierwszego stopnia dla trzech form zgrupowanego szeregu liczby kwiatów w koszykach bławatka.

Momenty	Formy szeregu		
	I	II	III
m'_{11}	—1.645	—1.526	—1.310
m_{11}	—4.935	—4.578	—3.930
m'_{12}	+ 1.382	+ 1.632	+ 1.913
m_{12}	+ 4.146	+ 5.739	+ 5.739

Błędy powodowane przez grupowanie wypadły tu daleko większe niż dla średniej artymetycznej. Nie jest to wcale dziwne wobec tego, że przy obliczaniu momentów pierwszego stopnia błędy pochodzące od minus i plus wariantów nie równoważą się.

Momenty pierwszego stopnia są bardzo cenne, gdyż dają wymiar odchyleń wariantów od średniej, określają stopień rozsiewu wariantów, jak się to mówi.

13. Momenty drugiego stopnia. Są one oczywiście zawsze dodatnie. Obliczenie ich można oprzeć na równaniu, wyprowadzonym w końcu ust. 10. Napiszmy je w formie

$$\sum_1^n (u_i - u)^2 = \sum_1^n (u_i - a)^2 - nb^2$$

Wystarczy to równanie podzielić przez n, by otrzymać odnośny podstawowy wzór

$$m_2 = \frac{\sum_1^n (u_i - a)^2}{n} - b^2$$

Przez wprowadzenia pojęcia pomocniczych momentów względem wartości wyjściowej powyższy wzór przyjmie formę: $m_2 = \mu_2 - \mu_1^2$

Dla szeregów rozdzielczych pomocniczy moment μ_2 oblicza się na podstawie wzoru:

$$\mu_2 = \frac{\sum_1^n f_i (x_i - a)^2}{n}$$

Obliczenie przeprowadza się według schematu, który przedstawię dla liczb kwiatów w koszykach sałatnicy (tab. 13,1). Została wzięta ta sama wartość $a = 17$.

Tabela 13,1

Liczby kwiatów w koszykach sałatnicy

x	f	$(x-a)^2$	$f(x-a)^2$
11	2	36	72
12	7	25	175
13	16	16	256
14	29	9	261
15	42	4	168
16	45	1	45
17	54	0	0
18	46	1	46
19	43	4	172
20	36	9	324
21	19	16	304
22	15	25	375
23	1	36	36
24	2	49	98
25	2	64	128
	359		2460

Otrzymujemy $\mu_2 = \frac{2460}{359} = 6.852$, $m_2 = 6.852 - 0.279^2 =$ $= 6.852 - 0.078 = 6.774$.

Z materiałem zgrupowanym w klasy obliczenia prowadzi się tak samo, traktując numery klas jako wartości wariantów. Otrzymany prowizoryczny wynik mnoży się przez k w a d r a t szerokości klas. Weźmiemy znowu materiał bławatka. Obliczenia przeprowadzimy osobno dla wszystkich trzech form grupowania i dla materiału niegrupowanego, tak samo jak to zrobiliśmy dla średniej arytmetycznej.

Dla pierwszej formy szeregu (tab. 13,2) otrzymujemy przy

$a = 10\frac{2}{3} : \mu'_2 = \frac{1409}{300} = 4,667,\ m'_2 = 4.697 - 0.937^2 = 4.697 -$

$- 0.878 = 3.819.$

Mnożąc przez kwadrat szerokości klas, otrzymamy ostatecznie $m_2 = 34.371$. Można przyjąć $m_2 = 34.37$.

Tabela 13,2

Liczby kwiatów w koszykach bławatka

Pierwszy sposób grupowania.

Zakresy klas	x'	f'	$(x' - a)^2$	$f'(x' - a)^2$
19—	6⅔	1	16	16
22—	7⅔	6	9	54
25—	8⅔	16	4	64
28—	9⅔	43	1	43
31—	10⅔	71	0	0
34—	11⅔	56	1	56
37—	12⅔	50	4	200
40—	13⅔	34	9	306
43—	14⅔	9	16	144
46—	15⅔	5	25	125
49—	16⅔	7	36	252
52—	17⅔	1	49	49
55—	18⅔	0	64	0
58—	19⅔	0	81	0
61—	20⅔	1	100	100
		300		1409

SPIS RZECZY:

Strona

Poprawki i uzupełnienia

Strona:	Wiersz:	Zamiast:	Ma być:
49	16 od góry	$\bar{u} = a + = \frac{\sum_1^l f_i(x_i - a)}{\sum_1^i f_i}$	$\bar{u} = a + \frac{\sum_1^l f_i(x_i - a)}{\sum_1^i f_i}$
50	13 „	$\div 2$	$\div 1$
53	4 „	$+0{,}40$	$+0{,}04$
55	7 od dołu	$m' = 14{,}5 + 1{,}223 = 14{,}673$, skąd $m' = 146{,}7$.	$m' = 14{,}45 + 1{,}223 = 15{,}673$, skąd $m' = 156{,}7$.
56	4 od góry	$\ldots = \sum_1^n \left(a u_i - \bar{u}\right)^2 + \ldots$	$\ldots = \sum_1^n \left(u_i - \bar{u}\right)^2 + \ldots$
80	4 „	śledzony	śledziony
80	5 „	śledzony	śledziony
91	4 „	Jaóba	Jakóba
102		$\overset{0}{\frac{1}{32}} 3 \times \overset{1}{\frac{1}{32}} 10 \times \ldots$	$\overset{0}{\frac{1}{32}} 5 \times \overset{1}{\frac{1}{32}} 10 \times \ldots$
110	10 od dołu	$\ldots np\left[q^{n-1} + \right] (n-1) \ldots$	$\ldots np\left[q^{n-1} \div (n-1) \ldots\right.$
117	1 „	$\ldots \frac{\xi^{2k+1}}{2^k\,!(2k+1)}$	$\ldots \frac{\xi^{2k+1}}{2^k k!\,(2k+1)}$
127	3 „	$\frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \ldots\right) + \frac{1}{120} + \ldots)$	$\frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \ldots + \frac{1}{120} + \ldots\right)$

Dla drugiej formy szeregu przy $a = 11$ wypada:

$$\mu'_2 = \frac{1315}{00} = 4.383, \quad m'_2 = 4.383 - 0.623^2 = 4.383 - 0.388 = .995,$$

$m_2 = 9 \times 3.995 = 35.995$. Można przyjąć $m_2 = 35.99$. Szczegółów obliczeń dla tej formy i dla trzeciej nie podaję, pozostawiając czytelnikowi przeprowadzenie rachunków.

Wreszcie dla trzeciej formy szeregu przy $a = 11^1/_3$ wypada:

$$\mu'_2 = \frac{1205}{300} = 4.017, \quad m'_2 = 4.017 - 0.243^2 = 4.017 - 0.059 = 3.958.$$

Stąd $m_2 = 35.622$. Można przyjąć $m_2 = 35.62$.

Niegrupowana forma omawianego materiału daje $m_2 = 34.704$. Wobec tego pierwsza forma szeregu dała wynik za mały, dwie pozostałe — za duży. Różnice wynoszą kolejno: -0.33, $+1{,}29$, $+0.92$. Są one większe niż dla średniej arytmetycznej, bo błędy pochodzące z minus i plus wariantów nie równoważą się.

Angielski statystyk S h e p p a r d zaproponował użycie poprawki, która częściowo wyrównuje błędy, jakie występują przy obliczeniu momentu drugiego stopnia skutkiem grupowania wariantów. Wynik wypada mianowicie z reguły za duży. Poprawkę przeto odejmuje się. Wynosi ona $-\frac{1}{12}\lambda^2$, gdzie λ jest szerokością klas. Ta poprawka nie zawsze poprawia wynik. Warunkiem powodzenia jest przede wszystkim stopniowość w zmianach częstości w przejściu od jednego wyrazu do drugiego w pierwotnej formie szeregu. Następnie forma szeregu powinna być nietylko symetryczna, ale jeszcze zbliżona do normalnej krzywej prawdopodobieństwa, o której będzie mowa w ust. 25.

W naszym przykładzie mamy szereg nie odbiegający zbytnio od normalnej krzywej prawdopodobieństwa, ale pierwotna jego forma wykazuje silne przeskoki w częstościach. To też poprawka S h e p p a r d a do pierwszej formy szeregu nie da się wogóle zastosować, bo wynik i tak wypadł za mały. Drobne

wyrównanie otrzymuje się tylko dla trzeciej formy szeregu, dla której po wprowadzeniu poprawki wypada wartość momentu równa 34.87.

Trzecia forma szeregu bławatka, ogólnie biorąc, wypada więc najlepsza, bo i średnia arytmetyczna była bliska prawdziwej. Była to forma szeregu najlepiej wyrównana. Można przyjąć zatem za prawidło ogólne, by wybierać formy szeregów możliwie najbardziej wyrównane. Można zalecić stosowanie do nich poprawki S h e p p a r d a, trzeba jednak zawsze mieć na uwadze podane powyżej zastrzeżenia.

Momenty drugiego stopnia mają większe znaczenie od wszystkich innych, a mianowicie o tyle, że oprócz charakterystyki materiału statystycznego, mającego prawie zawsze charakter próby, odgrywają one poważną rolę przy ocenie stopnia, w jakim dana próba charakteryzuje całość populacji. Będzie o tym mowa w rozdziale V.

14. Momenty trzeciego stopnia. Wstawiając do zasadniczego wzoru

$$m_3 = \frac{\sum_1^n (u_i - \bar{u})^3}{n}$$

$\bar{u} = a + b$, otrzymujemy wzór roboczy dla tego rodzaju momentów

$$m_3 = \frac{\sum_1^n (u_i - a - b)^3}{n} = \frac{\sum_1^n (u_i - a) - 3b \sum_1^n (u_i - a)^2 +}{n}$$

$$\frac{+ 3b^2 \Sigma (u_i - a) - \sum_1^n b^3}{n} = \mu_3 - 3b\mu_2 + 2b^3 \mu_3 - 3\mu_1 \mu_2 + 2\mu_1^3$$

Widzimy, że do obliczenia momentu trzeciego stopnia trzeba oprócz pomocniczego momentu μ_3 obliczyć jeszcze dwa poprzednie momenty tego rodzaju: μ_1 i μ_2.

Tabela 14,1

Liczby kwiatów w koszykach sałatnicy.

x	f	$(x-a)^3$	$f(x-a)^3$
11	2	—216	—432
12	7	—125	—875
13	16	—64	—1024
14	29	—27	—783
15	42	—8	—336
16	45	—1	—45
17	54	0	—3495
18	46	+ 1	46
19	43	+ 8	344
20	36	+ 27	972
21	19	+ 64	1216
22	15	+ 125	1875
23	1	+ 216	216
24	2	+ 343	686
25	2	+ 512	1024
	359		+ 6379

Dla szeregów rozdzielczych mamy wzór na pomocniczy moment μ_3:

$$\mu_3 = \sum_{1}^{l} \frac{f_i (x_i - a)^3}{n}$$

Według tego wzoru możemy tytułem przykładu przeprowadzić obliczenia dla liczby kwiatów w koszykach sałatnicy (tab. 14,1). Otrzymamy $\mu_3 = + \frac{2884}{359} = + 8.034$.

Poprzednio otrzymaliśmy:

$\mu_1 = 0.279$ i $\mu_2 = 6.852$. Wypadnie więc $3\mu_1\mu_2 = 5.735$ i $2\mu_1^3 =$ =0.044. Ostateczny wynik będzie:

$$m_3 = + 8.034 - 5.735 + 0.044 = 2.343.$$

Jeżeli warianty są zgrupowane w klasy, obliczenia przeprowadza się z początku tak samo. Potem mnoży się wynik przez trzecią potęgę szerokości klas. Jako przykład weźmiemy materiał bławatka zgrupowany trzecim sposobem (tab. 14,2). Ta forma szeregu wypadła bowiem najlepsza. Otrzymamy:

$$\mu_3 = + \frac{2497}{300} = + 8.323.$$

Poprzednio mieliśmy: $\mu'_1 = + 0.243$, $\mu'_2 = 4.017$. Wobec tego $m'_3 = 8.323 - 3 \times 0.243 \times 4.017 + 2 \times 0.243^3 = 8.323 - 2.928 + 0.029 = + 5.424$. Ten wynik jest wyrażony w przedziałach klasowych. Mnożąc go przez sześcian szerokości klas, to znaczy przez 27, otrzymamy ostatecznie $m_3 = 146.448$. Można przyjąć: $m_3 = 146.4$.

Tabela 14,2

Liczba kwiatów w koszykach bławatka.

Trzecia forma szeregu.

Zakres klas	x'	f'	$(x' - a)$	$f'(x' - a)^3$
21—	7 1/3	5	—64	—320
24—	8 1/3	11	—27	—297
27—	9 1/3	38	—8	—304
30—	10 1/3	61	—1	—61
33—	11 1/3	63	0	—982
36—	12 1/3	49	+1	49
39—	13 1/3	37	+8	296
42—	14 1/3	22	+27	594
45—	15 1/3	5	+64	320
46—	16 1/3	4	+125	500
51—	17 1/3	3	+216	648
54—	18 1/3	1	+343	343
57—	19 1/3	0	+512	0
60—	20 1/3	1	+729	729
		300		+3479

Dla tego materiału zgrupowanego pierwszym i drugim sposobem otrzymamy dla momentu trzeciego stopnia wartości 140.9 i 136.5. Przeprowadzenie rachunków pozostawiam czytelnikowi. Dokładna wartość wynosi $m_3 = 149.4$. Widzimy, że i w tym przypadku trzeci sposób grupowania jest najlepszy. Przy obliczaniu momentów trzeciego stopnia błędy wynikające z grupowania nie powinny być duże, gdyż wpływy minus i plus wariantów częściowo znoszą się. W naszym przykładzie te błędy są jednak znaczne, co wypływa z charakteru materiału, przez silne przeskoki w częstościach w pierwotnej jego formie. Przyjęte jest przy obliczaniu momentów trzeciego stopnia nie stosować żadnej poprawki na grupowanie.

Momenty trzeciego stopnia mogą mieć wartości dodatnie i ujemne. Stoi to w związku z charakterem skośności. Szeregi zupełnie symetryczne dałyby wartość równą zeru. Wartość absolutna wypada tym większa, im szereg jest bardziej skośny.

15. Momenty czwartego stopnia. Dla obliczenia tych momentów wstawiamy, jak zwykle, do wzoru zasadniczego

$$m_4 = \frac{\sum_1^n (u_i - \bar{u})^4}{n}$$

$\bar{u} = a + b$. Otrzymamy:

$$m_4 = \frac{\sum_1^n (u_i - a - b)^4}{n} = \frac{1}{n}\left[\sum_1^n (u_i - a)^4 - 4b\sum_1^n (u_i - a)^3 + \right.$$

$$+ 6b^2 \sum_1^n (u_i - a)^2 - 4b^3 \Sigma (u_i - a) + \sum_1^n b^4 = \mu_4 - 4b\mu_3 + 6b^2\mu_2 -$$

$$- 3b^4 = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_3 - 3\mu_1^4$$

Dla szeregów rozdzielczych mamy wzór roboczy:

$$\mu_4 = \frac{\sum_1^l f_i (x_i - a)^4}{n}$$

Według niego tytułem przykładu obliczymy moment czwartego stopnia dla liczby kwiatów w koszykach sałatnicy (tab. 15,1).

Tabela 15,1

Liczba kwiatów w koszykach sałatnicy.

x	f	$(x-a)^4$	$f(x-a)^4$
11	2	1296	2592
12	7	625	4375
13	16	256	4096
14	29	81	2349
15	42	16	672
16	45	1	45
17	54	0	0
18	46	1	46
19	43	16	688
20	36	81	2916
21	19	256	4864
22	15	625	9375
23	1	1296	1296
24	2	2401	4802
25	2	4096	8192
	359		46308

Będziemy mieli

$$\mu_4 = \frac{46308}{300} = 128{,}99$$

Poprzednio otrzymaliśmy: $\mu_1 = b = 0.279$ $\mu_2 = 6.852$ $\mu_3 = 8.034$.

Wobec tego otrzymamy:

$m_4 = 128.99 - 4 \times 0.279 \times 8.034 + 6 \times 0.279^2 \times 6.852 -$
$- 3 \times 0.279^4 = 128.99 - 8.95 + 3.19 - 0.02 = 123.2.$

Obliczanie momentów czwartego stopnia dla materiałów zgrupowanych w klasy przeprowadza się tak samo, traktując

numery klas jako wartości wariantów. Wynik otrzymuje się w przedziałach klasowych jako jednostce pomiarowej. Ostateczną wartość otrzymuje się, mnożąc przez czwartą potęgę szerokości klas.

Jako przykład przeprowadzimy rachunki dla bławatka według materiału zgrupowanego trzecim sposobem (tab. 15,2). Otrzymamy.

$\mu'_4 = \frac{21803}{300} = 72.98$, $m'_4 = 66.30$. Mnożąc go przez $3^4 = 81$ otrzymujemy $m_4 = 5370$.

Tabela 15,2

Liczby kwiatów w szczytowych koszykach bławatka.

Trzecia forma szeregu.

Zakres klas	x'	f'	$(x' - a)^4$	$f'(x' - a)^4$
21—	7⅓	5	256	1280
24—	8¼	11	81	891
27—	9⅓	38	16	608
30—	10⅓	61	1	61
33—	11⅓	63	0	0
36—	12⅓	49	1	49
39—	13⅓	37	16	592
42—	14⅓	22	81	1782
45—	15⅓	5	256	1280
48—	16⅓	4	625	2500
51—	17⅓	3	1296	3888
54—	18⅓	1	2401	2401
57—	19⅓	0	4096	0
60—	20⅓	1	6561	6561
		300		21893

Dla pierwszego i drugiego sposobu grupowania wypadają wartości 4981 i 5582. Dokładna wartość wynosi 5915. Najlepiej wypadła w tym przypadku druga forma grupowania, nie trzecia jak dotąd. Różnice są znaczne, bo niema wyrównania działaniem plus i minus wariantów.

Przy obliczaniu momentów czwartego stopnia używa się poprawki, która wynosi:

$$-\lambda^4\left(\frac{1}{2}m_2+\frac{1}{80}\right),$$

gdzie λ jest szerokość klas. Poprawkę trzeba odejmować. Zastosowanie jej dla trzech form szeregu daje wartości: 4828, 5423 i 5212. W naszym przypadku poprawka wywołuje pogorszenie wyników.

Momenty czwartego stopnia są naturalnie zawsze dodatnie. Znaczenie ich polega na tym, że określają one, w jakim stopniu warianty są skupione przy średniej arytmetycznej. Im bardziej warianty są skupione przy średniej, tym mniejszy wypada moment czwartego stopnia — mniejszy nie większy, jak by się można było spodziewać.

ROZDZIAŁ III.

POCHODNE CHARAKTERYSTYKI MATERIAŁÓW STATYSTYCZNYCH.

16. Wstępne uwagi. W rozdziale II rozpatrzyliśmy podstawowe charakterystyki szeregów rozdzielczych. Trzeba teraz wyzyskać je dla bezpośredniego, można powiedzieć poglądowego, przedstawienia osobliwości materiałów statystycznych.

Najbardziej poglądową charakterystyką materiałów statystycznych jest wartość modalna, wartość wariantów najczęstsza. Niestety, jak to było wyjaśnione w ust. 9, wyznaczenie jej jest możliwe tylko dla obfitych materiałów statystycznych.

Średnia arytmetyczna jest mniej poglądową charakterystyką. Jeżeli materiał układa się w szereg rozdzielczy mniej więcej symetryczny, różni się ona nieznacznie od wartości modalnej. Dla szeregów jednobocznych mówi niewiele. Tym nie mniej jest to charakterystyka podstawowa, gdyż daje się zawsze ustalić i nadto służy za podstawę do obliczenia innych charakterystyk.

Bardzo cenne są częściowe momenty pierwszego stopnia, gdyż dają bezpośredni wyraz rozsiewu wariantów wokół średniej arytmetycznej.

Wszystko to jednak nie wystarcza i trzeba wprowadzić jeszcze inne charakterystyki. Będą to funkcje charakterystyk podstawowych i przeto mogą być nazwane *p o c h o d n y m i*. Do rozpatrzenia ich teraz przystąpimy.

17. Przeciętne i średnie odchylenie. Trzeba przede wszystkim ustalić charakterystykę, która by dawała obraz stopnia zmienności wariantów, obraz stopnia ich rozsiewu. Zadanie to spełniają częściowo momenty pierwszego stopnia. Jest to jednak niedostateczne o tyle, że jest ich dwa: osobny dla minus wariantów, osobny dla plus wariantów. Trzeba mieć jeszcze ogólną charakterystykę rozsiewu, wspólną dla wszystkich wariantów.

Do tego służyć może przeciętne odchylenie wariantów od ich średniej arytmetycznej, nazywane w skrócie odchyleniem przeciętnym. Jest to średnia arytmetyczna absolutnych wartości wszystkich odchyleń wariantów od średniej arytmetycznej. Można ją zatem określić przez równanie:

$$\vartheta = \frac{\sum\limits_{1}^{n} \left| u_i - \bar{u} \right|}{n}$$

Obliczenie tej wielkości przeprowadza się za pomocą częściowych momentów pierwszego stopnia, gdyż jak to widzieliśmy w ust. 12 jeden z nich określa średnią wielkość odchyleń ujemnych, drugi — dodatnich. Trzeba wziąć średnią ważoną ich absolutnych wartości.

$$\vartheta = \frac{|m_{11}|\, n_1 + m_{12}\, n_2}{n_1 + n_2}$$

gdzie n_1 jest liczbą wariantów mniejszych od średniej, n_2 liczbą wariantów większych od niej. Średnia musi być ważona, gdyż te częściowe momenty nie są równoważne, reprezentując na ogół różne liczby wariantów.

Zestawiając wzory dla obu momentów, otrzymuje się dla wykonania obliczeń równanie:

$$\vartheta = \frac{\sum\limits_{1}^{k} f_i (a - x_i) + \sum\limits_{K+1}^{l} f_i (x_i - a) + (n_1 + n_2) b}{n}$$

Obliczenie łatwo jest przeprowadzić, wyzyskując liczby rachunków do obliczenia średniej arytmetycznej. I tak dla liczby kwiatów w koszykach sałatnicy, gdzie jest 195 minus wariantów i 164 plus wariantów, będziemy mieli:

$$\vartheta = \frac{327 + 427 + (195-164) \times 0.279}{359} =$$

$$= \frac{327 + 427 + 31 \times 0.279}{359} = \frac{762.65}{395} = 2.124$$

Dla materiałów zgrupowanych rachunki przeprowadza się tak samo, operując numerami klas jak wartościami wariantów. Ostateczny wynik otrzymuje się, mnożąc przez szerokość klas. Dla trzech form materiału bławatka wypadają wartości przeciętnego odchylenia równe 4.993, 4.731 i 4.665. Dokładna wartość wynosi 4.623. Trzecia forma grupowania daje zatem wynik najlepszy, tak jak dla wszystkich innych charakterystyk z wyjątkiem momentu czwartego stopnia.

Odchylenie przeciętne jest najlepszym, bo najbardziej poglądowym, przedstawieniem wielkości odchyleń od średniej, a przeto i zmienności wariantów. Używając tej wielkości, należy jednocześnie podawać oba częściowe momenty pierwszego stopnia, charakteryzujące osobno odchylenia ujemne i dodatnie. Jako przykład można podać przeliczenia S m o s a r s k i e g o [1]) dla średnich dobowych temperatur Poznania w poszczególnych miesiącach lat 1848 — 1922 (tab. 17,1). Ta tabela zawiera nadto średnie arytmetyczne temperatur (*m*) i wielkości σ i $\frac{\vartheta}{\sigma}$, o których będzie mowa później.

Pomimo swoich zalet odchylenie przeciętne rzadko jest obliczane. Jedynie meteorologowie korzystają stale z tej wielkości, przykładem przytoczona praca S m o s a r s k i e g o.

---

[1]) W. S m o s a r s k i. Temperatura i opady w Wielkopolsce. — Prace Komisji Matematyczno - Przyrodniczej Tow. Przyjaciół Nauk w Poznaniu. Seria A. Tom 2 (1925).

Zwykle dla charakterystyki zmienności wariantów bierze się t. zw. odchylenie średnie, które jest pierwiastkiem kwadratowym z momentu drugiego stopnia: $\sigma = \sqrt{m_2}$.

Oczywiście ta wielkość nie daje tak dobrego obrazu wielkości odchyleń, jak odchylenie przeciętne. W niektórych przypadkach daje wręcz złe o tym pojęcie, mianowicie kiedy wariantów jest mało i trafią się wśród nich niektóre wyjątkowo silnie odbiegające od średniej arytmetycznej. Wtedy te ostatnie warianty dają wyjątkowo duże odchylenia, które będąc podniesione do kwadratu przy obliczeniu średniego odchylenia wpływają nadmiernie na jego wartość. Przy obfitych materiałach tak źle nie jest, bo wprawdzie mogą się w nich trafić takie wyjątkowe warianty, ale ich liczba w stosunku do ogólnej ilości wariantów nigdy nie jest duża.

Jeżeli pomimo swojej mniejszej poglądowości średnie odchylenie jest powszechnie używane zamiast przeciętnego, tłumaczy się to głównie tym, że przeciętne odchylenie poza przedstawieniem rozsiewu wariantów nie znajduje żadnego prawie innego zastosowania, podczas gdy odchylenie średnie znajduje bardzo ważne zastosowanie w ocenie wiarogodności charakterystyk. Jak to będzie dokładniej objaśnione w rozdziale V, materiały statystyczne prawie zawsze są próbami wziętymi z jakiejś populacji. Przeto obliczone na ich podstawie charakterystyki tylko w przybliżeniu określają osobliwości populacji. Określenie, w jakim stopniu to czynią, wymaga znajomości średniego odchylenia.

Ta okoliczność jeszcze nie wystarczyłaby do porzucenia przeciętnego odchylenia, ale do tego dochodzi jeszcze ta właściwość obu omawianych odchyleń, że stosunek między nimi jest mało zmienny. Jak to widać z tablicy 17,1, odchylenie przeciętne jest mniejsze od średniego, ale stosunek jednego do drugiego waha się dla odnośnego materiału w granicach od 0.755 do 0.858. W innych przypadkach jest mniej więcej tak samo. W ten sposób, mając tylko odchylenie średnie, można mieć przybliżone pojęcie o rozsiewie wariantów.

Tabela 7,1

Średnia dobowa temperatur powietrza w Poznaniu dla poszczególnych miesięcy i lat okresu 1848 — 1922.

Charakterystyki	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Rok
m	—2.0	—0.7	+2.4	7.8	13.1	17.1	18.6	17.6	13.7	8.4	2.6	—0.6	8.2
m_{11}	—2.53	—2.74	—2.07	—1.42	—1.41	—1.21	—1.30	—0.95	—0.95	—1.32	—1.64	—2.34	—0.62
m_{12}	2.13	1.86	1.79	1.39	1.44	1.22	1.04	1.19	1.04	1.42	1.50	1.36	0.58
ϑ	2.32	2.22	1.92	1.40	1.43	1.22	1.15	1.06	0.99	1.37	1.57	2.06	0.60
σ	3.01	2.94	2.32	1.75	1.77	1.50	1.34	1.28	1.27	1.67	1.90	2.53	0.70
$\frac{\vartheta}{\sigma}$	0.770	0.775	0.828	0.800	0.808	0.813	0.858	0.828	0.780	0.820	0.826	0.814	0.790

Do tego dochodzi jeszcze jedno: odchylenie przeciętne, lepiej charakteryzując poszczególne próby, nieco gorzej od średniego charakteryzuje populacje, z której próby były wzięte. (por. ust. 30). Wreszcie wzór dla średniego odchylenia jest prostszy niż dla przeciętnego.

Dla omawianych stale przykładów sałatnicy i bławatka średnie odchylenie wynosi: dla pierwszego z tych materiałów $\sigma = \sqrt{6.774} = 2.603$, dla drugiego użytego bez grupowania $\sigma = \sqrt{34.70} = 5.891$. Sposób grupowania wpływa naturalnie na otrzymywany wynik. Dla trzech form szeregu bławatka mamy wartości $\sqrt{33.621} = 5.798$, $\sqrt{35.245} = 5.937$ i $\sqrt{34.872} = 5.905$.

Stosunek odchylenia przeciętnego do średniego wypada dla sałatnicy 0.815 a dla bławatka 0.785. Są to liczby, które mieszczą się w granicach wartości, jakie dał materiał meteorologiczny Smosarskiego. Dla szeregów rozdzielczych formy normalnej, której istota będzie wyjaśniona w ust. 25, ten stosunek równa się $\sqrt{\frac{2}{\pi}} = 0.798$. Jest to tzw. wzór Cornu.

Warto jest jeszcze poświęcić trochę uwagi nazwom omawianych w tym ustępie charakterystyk. Panuje tu wielkie zamieszanie. W języku polskim „przeciętny" znaczy to samo co „średni". Należałoby zmienić te terminy. Najlepiej byłoby dla uniknięcia nieporozumień nazywać przeciętne odchylenie „średnim liniowym" a średnie „średnim kwadratowym" tak jak to czynią Francuzi (deviation linéaire moyenne et quadratique moyenne). W języku angielskich twórców współczesnej statystyki tych niejasności niema: tam odchylenie przeciętne nazywa się mean deviation, średnie zaś standart deviation.

Na ostatku warto jest zaznaczyć, że rozsiew wariantów może być scharakteryzowany bez względu na ich średnią arytmetyczną. Według propozycji jednego z włoskich statystyków można do tego celu użyć średniej arytmetycznej absolutnych wartości różnic między wariantami, zestawiając je we wszelki możliwy sposób. Obliczenie takiej wielkości nie jest zawiłe, chociaż mozolne. Wątpliwe jest tylko, czy taka charakterystyka jest lepsza od omówionych w tym rozdziale charakterystyk zmienności wariantów.

18. Wskaźnik zmienności. Dla lepszego zobrazowania zmienności materiału statystycznego dobrze jest obliczyć w s k a ź n i k z m i e n n o ś c i. J e s t t o s t o s u n e k m i a r y o d c h y l e ń o d ś r e d n i e j a r y t m e t y c z n e j. Stosownie do tego, czy się weźmie za taką miarę odchylenie przeciętne czy średnie, będziemy mieli 2 równania:

$$v_\vartheta = \frac{\vartheta}{m} \text{ i } v_\sigma = \frac{\sigma}{m}$$

Dla liczby kwiatów w koszykach sałatnicy otrzymamy:

$$v_\vartheta = \frac{2.12}{17.28} = 0.123, \quad v_\sigma = \frac{2.60}{17.28} = 0.150.$$

Dla bławatka zaś biorąc dokładne wartości charakterystyk:

$$v_\vartheta = \frac{4.263}{34.77} = 0.133 \text{ i } v_\sigma = \frac{5.891}{34.77} = 0.169.$$

Zwykle wyraża się wskaźnik zmienności w procentach, a więc dla sałatnicy będą wartości 12.3% i 15.0% a dla bławatka 13.3% i 16.9%.

Wobec powszechnego użycia odchylenia średniego podawane w literaturze wartości wskaźnika zmienności prawie zawsze odnoszą się do wielkości v_σ. Jako przykład można przytoczyć następujące dane co do cech człowieka według źródeł angielskich (tab. 18,1). Mamy tu wartości bardzo różne

Tabela 18,1

Wskaźnik zmienności v_σ dla cech człowieka w procentach.

Cechy	Mężczyźni	Kobiety
Waga śledzony u zdrowych	38.2	—
Waga śledzony u chorych	50.6	—
Waga serca u chorych	38.4	—
Waga ciała	10.1	13.4
Siła dłoni	14.1	18.6
Waga mózgu	9.2	9.7
Wzrost	4.0	4.1
Obwód czaszki	2.9	2.7

Przykładów wielkości v_{ϑ} trzeba szukać u meteorologów. Dla przytoczonych w poprzednim ustępie danych Smosarskiego o temperaturze powietrza nie można obliczyć wskaźnika zmienności, chybaby się wzięło temperatury absolutne za przykładem Anglików. Ale znajdujemy w tej samej rozprawie dane o opadach w Poznaniu za ten sam okres 1848—1922 (tab. 18.2). Mamy tam tylko odchylenia przeciętne i obliczone na ich podstawie wskaźniki zmienności dla poszczególnych miesięcy i dla całego roku. Warto jest zwrócić uwagę na to, o ile mniejszy jest ten wskaźnik dla roku w porównaniu do miesięcy.

19. Wskaźnik skośności. Z kolei trzeba zająć się charakterystyką skośności szeregów rozdzielczych. Jest to zadanie dosyć zawiłe i może być rozwiązane w dwojaki sposób.

Pierwszy sposób opiera się na rozkładzie częstości względem wyrazu o największej częstości, reprezentującego wartość modalną zmiennej ewentualnej. Jeżeli wyrazy jednakowo odległe od wspomnianego wyrazu porównamy ze sobą, ujawnia się, że ich częstości prawie zawsze są nierówne. Jeżeli wyrazy o większej wartości zmiennej ewentualnej mają częstość więk-

szą od wyrazów mniejszej wartości tej zmiennej, uważać można skośność za dodatnią, w przeciwnym razie za ujemną. Tak kwestia skośności szeregów rozdzielczych była traktowana dotychczas (por. ust. 6).

Takie porównania musiałyby dotyczyć wszystkich wyrazów szeregu, co jest niewygodne. Otóż można je zastąpić przez jedno tylko porównanie — wartości modalnej ze średnią arytmetyczną. Łatwo jest bowiem zdać sobie sprawę, że przy skośności dodatniej średnia jest większa od wartości modalnej,

Tabela 18,2

Opady atmosferyczne w Poznaniu w okresie 1848—1922 (w mm)

Charakterystyki	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Rok
m	32	26	32	36	46	57	70	62	43	35	33	36	508
ϑ	13	14	14	16	19	24	31	27	20	17	15	16	63
v_σ	41	54	54	44	41	42	45	44	46	48	46	43	12

przy skośności ujemnej natomiast mniejsza od niej. Przy tym odchylenie średniej od modalnej jest tym większe, im skośność jest silniejsza. W ten sposób powyższe odchylenie można uważać za miarę skośności.

Samo odchylenie od modalnej nie wystarcza jednak dla charakterystyki skośności. Trzeba jeszcze wziąć pod uwagę rozsiew wariantów. Istotnie dwa szeregi jednakowo skośne, to znaczy o jednakowych różnicach częstości w jednakowo od modalnej odległych wyrazach wykażą różne odchylenie średniej modalnej wartości, jeżeli rozsiew wariantów jest różny: przy większym rozsiewie rozpatrywane odchylenie będzie większe,

przy mniejszym — mniejsze. Można coprawda rozumować tak tylko w odniesieniu do wyrazów szeregu niezbyt daleko odległych od wartości modalnej, ale to nie zmienia zasadniczo istoty zagadnienia. Wynika z tego rozumowania, że aby mieć miarę skośności, trzeba podzielić odchylenie średniej od modalnej przez miarę rozsiewu wariantów, przez przeciętne albo średnie odchylenie. Na tej zasadzie Pearson zaproponował jako miarę skośności funkcję:

$$\alpha = \frac{m - A}{\sigma}$$

gdzie A jest wartością modalną. Tę funkcję będziemy nazywali wskaźnikiem skośności pierwszego rodzaju.

Dla zobrazowania zagadnienia rozpatrzmy kilka przykładów. Pierwszym będzie liczba kwiatów w koszykach sałatnicy. Dla niej mieliśmy $m = 17.279$, $A = 17$, $\sigma = 2.603$. Wobec tego wypada $\alpha = +0.107$. Jest to słaba skośność dodatnia.

Następnie weźmiemy liczbę wewnętrznych listków okrywy u tej samej rośliny (tab. 6,2). Obliczamy średnią i średnie odchylenie. Wypada $m = 8.1967$, $\sigma = 0.6094$. Ponieważ $A = 8$, otrzymujemy $\alpha = +0.323$. Skośność jest tak samo dodatnia, ale silniejsza.

Wreszcie rozpatrzymy liczbę kwiatów języczkowych w szczytowych koszykach *Senecio fluviatilis* (tab. 6,4). Dla niego $m = 7.6114$, $A = 8$, $\sigma = 0.7155$. Wobec tego $\alpha = -0.543$. Wskaźnik skośności zgodnie z charakterem szeregu wypadł ujemny.

Powyższy sposób ujęcia zagadnienia skośności jest prosty i jasny. Niestety, nie zawsze można je tak traktować, gdyż wyznaczenie wartości modalnej jest trudne dla szeregów ze zgrupowanymi wartościami wariantów i niemożliwe dla materiałów mało licznych. Dla zaradzenia złemu trzeba szukać innej drogi, mianowicie trzeba wziąć za punkt wyjścia moment trzeciego stopnia. Jest on równy zeru dla szeregów dokładnie symetrycznych. Dla szeregów skośnych może wypaść dodatni

lub ujemny zależnie od rozkładu częstości. Trzeba przy tym naturalnie wyeliminować, tak samo jak poprzednio, stopień rozsiewu wariantów.

W ten sposób dochodzi się do wskaźnika skośności drugiego rodzaju określonego równaniem

$$\gamma_1 = \frac{m_3}{\sigma^3}$$

Nietrudno jest wykazać, że jest to osobliwego rodzaju moment trzeciego stopnia. W istocie wstawiamy w powyższe równanie wzór momentu trzeciego stopnia

$$m_3 = \frac{\sum_1^n (u_i - \bar{u})^3}{n}$$

Otrzymamy

$$\gamma_1 = \frac{\sum_1^n (u_i - \bar{u})^3}{n\,\sigma^3}$$

Ponieważ średnie odchylenie jest wielkością stałą dla danego szeregu, można je wprowadzić pod znak sumy:

$$\gamma_1 = \frac{\sum_1^n \left(\frac{u_i}{\sigma} - \frac{\bar{u}}{\sigma}\right)^3}{n}$$

Widzimy, że jest to takie same równanie jak powyżej przytoczone dla momentu trzeciego stopnia, tylko wymiar wielkości wariantów jest inny: za jednostkę pomiarową jest przyjęte średnie odchylenie. Jest to zatem osobliwy moment trzeciego stopnia, który można nazwać momentem trzeciego stopnia drugiego rodzaju.

Zobaczmy, jakie wartości przybiera ten nowy wskaźnik skośności w przykładach poprzednio omówionych. Dla liczby kwiatów w koszykach sałatnicy mieliśmy $m_3 = +2.343$ i przeto wypada $\gamma_1 = +0.133$. Dla listków okrywy tej samej rośliny

otrzymuje się $m_3 = +0.3900$ i zatem $\gamma_1 = +1.724$. Wreszcie dla liczby kwiatów języczkowych u *Senecio fluviatilis* $m_3 = -0.3732$ i $\gamma_1 = -1.019$.

Otrzymaliśmy wartości odmienne od wartości wskaźnika pierwszego rodzaju. Wyjątkowo może się zdarzyć, że nawet znak odwrotny. Tak jest dla materiału otrzymanego dla kwiatów języczkowych w szczytowych koszykach syberyjskiej rośliny *Senecio octoglossus* w pewnej kulturze Ogrodu Botanicznego w Dublanach (tab. 19,1). Wypadło tam: $m = 7.9802$, $\sigma = 0.4445$, $A = 8$, $m_3 = +0.00806$. Stąd wynika, że $\alpha = -0.445$, $\gamma_1 = +0.0918$.

Tabela 19,1

Liczby kwiatów języczkowych w koszykach
Senecio octoglossus

Liczby kwiatów:	5	6	7	8	9	10	11
Ich częstości:	1	2	14	277	4	4	1

Te niezgodności między dwoma rodzajami wskaźnika skośności pochodzą stąd, że pierwszy z nich określa różnice rozsiewu wariantów względem wartości modalnej, drugi zaś względem średniej arytmetycznej. Wskaźnik drugiego rodzaju nie daje tak jasnego obrazu skośności jak pierwszy, zato można go zawsze obliczyć, gdyż wartość modalna jest tu niepotrzebna a momenty zawsze dadzą się obliczyć.

Dla materiałów zgrupowanych w klasy wskaźnik drugiego rodzaju oblicza się tak samo jak dla materiałów niegrupowanych. Do tego nie potrzeba ostatecznych wartości charakterystyk m_3 i σ, wystarczą wartości prowizoryczne m'_3 i σ' wyrażone w szerokości klas. Przytem przy obliczeniu średniego odchylenia trzeba stosować poprawkę Shepparda w wymiarze $\frac{1}{12} = 0.0833...$ a nie w wymiarze $\frac{\lambda^2}{12}$.

Wykonując tytułem przykładu obliczenia dla trzeciej formy szeregu bławatka, otrzymamy: $m'_3 = +5.424$, $\sigma' = \sqrt{3.958 - 0.083} = \sqrt{3.875} = 1.968$ i wreszcie $\gamma_1 = +0.711$. Dokładna wartość wynosi: $\gamma_1 = +0.731$.

20. Wskaźnik skupienia wariantów i eksces. Ostatnią cechą szeregów rozdzielczych jest stopień skupienia wariantów w pobliżu wartości modalnej. Tu znowu wobec trudności w wyznaczaniu tej wartości trzeba zmienić punkt wyjścia i traktować zagadnienie jako kwestię skupienia wariantów przy średniej arytmetycznej. Wtedy można ją rozwiązać za pomocą momentu czwartego stopnia. Podobnie jak z momentu trzeciego stopnia można wyprowadzić miarę skośności, tak z momentu czwartego stopnia wywodzi się wskaźnik skupienia wariantów, określony przez równanie:

$$\beta = \frac{m_4}{\sigma^4}$$

Jest to moment czwartego stopnia drugiego rodzaju, gdyż można napisać:

$$\beta = \frac{\sum_1^n (u_i - \bar{u})^4}{n\,\sigma^4} = \frac{\sum_1^n \left(\frac{u_i}{\sigma} - \frac{\bar{u}}{\sigma}\right)^4}{n}$$

Przybiera on tym większe wartości, im bardziej skupione są warianty przy średniej. Moment czwartego stopnia wprawdzie zmniejsza się wtedy, ale średnie odchylenie zmniejsza się także i to w takim stopniu, że skutkiem podniesienia do czwartej potęgi zmniejszenie odchylenia wpływa silniej na wynik obliczenia aniżeli zmniejszenie momentu. Nietrudno jest to wykazać na przykładach. Weźmiemy w tym celu następujące szeregi rozdzielcze.

Rozpoczniemy od szeregów wklęsłych, w których warianty są skupione nie w środkowej części szeregu rozdzielczego, lecz na obu jego końcach. Wypadną przeto szczególnie małe wartości wskaźnika skupienia wariantów. Naprzykład za-

chmurzenie w Bazylei (tab. 6,14) da $m_4 = 0.03935$, $\sigma = 0.3922$ a więc

$$\beta = \frac{0.05935}{0.02365} = 1.664$$

Dalej liczby kwiatów w koszykach sałatnicy (tab. 6,2) będziemy mieli: $m_4 = 123.2$, $\sigma = 2.603$ i przeto $\beta = 2.685$.

Następnie dla liczby kwiatów języczkowych w koszykach *Senecio fluviatilis* (tab. 6,4) otrzymamy: $m_4 = 1.0385$ i $\sigma = 0.7155$ a więc $\beta = 3.963$.

Dla listków okrywy sałatnicy (tab. 6,2) wypadnie $m = 1.1754$, $\sigma = 0.60936$ i $\beta = 8.525$.

Dla liczby promieni meduzy *Pseudoclytia pentata* (tab. 6,5) otrzymuje się: $m_4 = 0.50818$, $\sigma = 0.4516$ i $\beta = 13.40$.

Wreszcie przytoczony w poprzednim ustępie *Senecio octoglossus* daje: $m_4 = 0.8209$, $\sigma = 0.4445$ i $\beta = 21.02$.

Widzimy, że wskaźnik skupienia wariantów waha się w daleko szerszych granicach niż wskaźnik skośności.

Omawianą właściwość szeregów rozdzielczych można także określić za pomocą innej jeszcze charakterystyki, zwanej ekscesem, oznaczanej symbolem γ_2. Eksces jest to wskaźnik skupienia wariantów pomniejszony o 3. Użycie tej charakterystyki ma swój sens przy porównywaniu dwubocznych szeregów rozdzielczych z normalną funkcją prawdopodobieństwa, o której będzie mowa w ust. 25. Dla niej wskaźnik skupienia wariantów wynosi 3. Skutkiem tego eksces wskazuje, o ile skupienie wariantów przy średniej jest mniejsze lub większe niż w szeregach z rozkładem częstości według normalnej funkcji prawdopodobieństwa. Skupienie słabsze niż u takich szeregów jest rzadkie i przeto eksces jest przeważnie dodatni. W naszych przykładach tylko dwa pierwsze mają eksces ujemny: zachmurzenie —1.336 i kwiaty sałatnicy —0.315. Poza tym jest on dodatni: od + 0.963 u *Senecio fluviatilis do* + 18.02 u *S. octoglossus.*

Dla materiałów zgrupowanych w klasy wskaźnik skupienia wariantów oblicza się podobnie jak wskaźnik γ_1, za pomocą

charakterystyk wyrażonych w szerokości klas. I tak dla trzeciej formy szeregu bławatka będziemy mieli: $m'_4 = 64.35$, $\sigma' = 1.968$, a zatem $\beta = 4.286$. Dokładna wartość wynosi

$$\beta = \frac{5332}{5.891^4} = 4.427$$

Eksces będzie miał odpowiednio wartości $+1.286$ i $+1.427$.

ROZDZIAŁ IV.

NIEZBĘDNE DLA STATYSTYKI WIADOMOŚCI Z NAUKI O PRAWDOPODOBIEŃSTWIE.

21. Pojęcie prawdopodobieństwa. Zagadnienie prawdopodobieństwa dotyczy częstości zdarzeń, powodowanych przez dwa różne rodzaje przyczyn. Są to z jednej strony przyczyny dające się dokładnie ustalić — przyczyny określone, z drugiej zaś strony takie, które można scharakteryzować tylko ogólnikowo — przyczyny nieokreślone. Naprzykład częstość wystąpienia czarnej karty w ciągnieniach z dobrze wymieszanej talii kart zależy od dającego się ustalić składu danej talii — od stosunku liczby czarnych kart do ogólnej ich ilości. Zależy nadto od niedającego się przewidzieć układu ich w talii i od niedających się przewidzieć tak samo ruchów ręki. Naturalnie zakłada się przytem, że szulerskie chwyty nie są stosowane, w przeciwnym bowiem razie będą działały tylko przyczyny określone i o prawdopodobieństwie nie będzie mowy.

Częstość zdarzeń, zgodnie z wywodami ust. 5, może być rozumiana dwojako — jako częstość bezwzględna, podająca liczbę danego rodzaju zdarzeń, i jako częstość względna, określona stosunkiem liczby zdarzeń do liczby przypadków, w których takie zdarzenia mogą wystąpić. W rozważaniach na temat prawdopodobieństwa chodzi zawsze o częstość względną, którą też dla skrócenia nazywa się zwykle po prostu częstością. Tak rozumiana częstość wyraża się liczbą mniejszą od jedności. W krańcowych przypadkach równa się ona zeru i jedności: równa się zeru — jeżeli zdarzenie nie występuje wcale, jedności — jeżeli wystąpi w każdym możliwym dla niego przypadku.

Częstości tego samego rodzaju zdarzeń wypadają różnie. Naprzykład francuski zoolog Buffon w r. 1777 określał częstość występowania orła w grze w orła i reszkę przy dużej ilości rzutów. Było ich 4040. Orzeł wystąpił 2048 razy, częstość

zatem wypadła 0.5069. Na początku XIX wieku angielski matematyk de Morgan powtórzył te zliczenia i otrzymał tę samą liczbę wystąpień orła na 4092 rzuty. Częstość otrzymał zatem równą 0.5005. Belgijski astronom Quetelet, o którym była mowa w ust. 2 jako o jednym z twórców współczesnej statystyki, także przeprowadził w r. 1846 podobne zliczenia, wyjmując kule z urny, zawierającej 20 kul białych i 20 czarnych. Takie ciągnienia oczywiście dają wyniki podobne do rzutów monetą. Biała kula wyszła 2066 razy na 4096 ciągnień, czyli częstość wypadła 0.5044.

Przytoczone częstości podobnych zdarzeń są zbliżone, nie są jednak równe. Tylko w bardzo rzadkich przypadkach wypadają one dokładnie sobie równe, ale to nie zmienia postaci rzeczy. Otóż fakt, że otrzymywane w doświadczeniu różne częstości dla tego samego rodzaju zdarzeń są zbliżone, jest powodowany przez działanie przyczyn określonych — przez symetrię monety w grze w orła i reszkę, przez równą ilość czarnych i białych kul w urnie itp. Różnice zaś między poszczególnymi wartościami częstości pochodzą z działania przyczyn nieokreślonych, jak ruchy ręki, podskoki monety przy uderzeniu o stół, układ kul w urnie i inne podobne okoliczności.

Gdyby przyczyny nieokreślone nie działały, częstość danego rodzaju zdarzeń wypadałaby zawsze jednakowa. Taką stałą wartość częstości nazywamy prawdopodobieństwem.

Faktycznie obserwowane częstości są, poza wyjątkowymi przypadkami, zbliżone do wartości prawdopodobieństwa. Jest ona w przytoczonych powyżej przykładach, jak się zaraz przekonamy, równa $\frac{1}{2}$. Zatem odchylenia częstości od niej w tych przykładach wynoszą 0.0069, 0.0005 i 0.0044.

Wartość prawdopobieństwa można obliczyć, jeżeli się zna dokładnie przyczyny określone. Wtedy, zakładając że przyczyny nieokreślone nie działają, trzeba najpierw ustalić liczbę jednakowo możliwych a zarazem różnych ewentualności, w których może wystąpić dane zdarze-

nie. Trzeba dalej zliczyć z pośród tych ewentualności takie, w których zdarzenie wystąpić musi. Otrzyma się w ten sposób dwie liczby całkowite. Dzieląc drugą przez pierwszą, otrzymamy wartość liczbową prawdopodobieństwa dla danego rodzaju zdarzeń.

I tak w grze w orła i reszkę są dwie ewentualności: moneta spadnie ukazując orła albo reszkę. Obie są równie możliwe, jeżeli moneta jest symetryczna i nieuszkodzona. Jedna z tych ewentualności daje rozpatrywane zdarzenie, jakim jest ułożenie się monety na stole orłem do góry. Prawdopodobieństwo wypada zatem równe $\frac{1}{2}$. Będzie ono inne, jeżeli skutkiem licznych uderzeń o stół moneta zostanie wykrzywiona lub w inny sposób uszkodzona.

W przypadku urny z 20 białymi i 20 czarnymi kulami mamy 40 ewentualności, bo wyjęcie każdej kuli jest jednakowo możliwe, jeżeli są one jednakowej wielkości i są wykonane z tego samego materiału. Z tych 40 ewentualności wystąpienie białej kuli jest pewne w 20, mianowicie kiedy ręka sięgająca do urny natrafi kulę tej barwy. Prawdopodobieństwo takiego zdarzenia jest $\frac{20}{40} = \frac{1}{2}$.

Podobnie jest z ciągnieniem kart z talii. Wyciągnięcie każdej karty jest jednakowo możliwe. Zatem ilość różnych możliwych ewentualności przy użyciu pełnej talii bez jockera jest równa 52. Jeżeli będzie chodziło o czarne karty, to tak samo wyciągnięcie każdej z nich jest równe możliwe — ilość odnośnych ewentualności wyniesie przeto 26. Prawdopodobieństwo wypadnie $\frac{26}{52} = \frac{1}{2}$. O ile by jednak chodziło o wyciągnięcie nie byle jakiej czarnej karty, lecz czarnego asa, to liczba ewentualności dla takiego zdarzenia spadłaby do 2 a prawdopodobieństwo do $\frac{2}{52} = \frac{1}{26}$.

Różnica między prawdopodobieństwem a faktycznie obserwowanymi częstościami na ogół maleje w miarę zwiększe-

nia liczby przypadków. Jest to tzw. *prawo wielkich liczb*, podstawowe w nauce o prawdopodobieństwie.

Matematyczne opracowanie tego prawa jest zasługą szwajcarskiego matematyka *Jaóba Bernoulli* (1654 — 1705), który strawił nad tym zagadnieniem 20 lat życia. Da się to prawo sformułować w sposób następujący:

Jeżeli w szeregu przypadków, w których występuje jakieś zdarzenie, prawdopodobieństwo jego wystąpienia nie ulegnie zmianie, to przy dostatecznie wielkiej ilości przypadków z prawdopodobieństwem dowolnie bliskim jedności, to znaczy dowolnie bliskim pewności, można spodziewać się, że częstość danego zdarzenia będzie dowolnie mało różniła się od jego prawdopodobieństwa.

W tej formie prawo wielkich liczb zostało przez *Bernoulli'ego* dowiedzione w drodze rozumowań matematycznych, których nie przytaczam z powodu ich zawiłości. Dla przyrodnika i wogóle dla badacza faktów jest to niedostateczne. Ponieważ to prawo dotyczy obserwacyj rzeczywistości, powinno być sprawdzone przez porównania z faktami. W tym celu trzeba je ująć inaczej, a mianowicie w sposób następujący.

Jeżeli będziemy brali coraz większe zbiory przypadków, w których występuje jakieś zdarzenie, to ze zwiększeniem tych zbiorów ale ciągle w tych samych warunkach coraz częściej wystąpią częstości tego zdarzenia, które coraz mniej będą się różniły od jego teoretycznie obliczonego prawdopodobieństwa.

Praktycznie rzecz biorąc, można porównanie tak ujętego prawa wielkich liczb z faktami przeprowadzić w sposób następujący. Zliczamy, ile razy wystąpi dane zdarzenie w zbiorze dziesięciu przypadków. Powtarzamy takie zliczenie dla wielkiej liczby, powiedzmy stu, innych zbiorów, liczących także po dziesięć przypadków. Otrzymamy na ogół różne częstości dla

różnych zbiorów. Obliczmy, ile każda częstość różni się od prawdopodobieństwa i jaka wypadnie średnia tych różnic. Następnie wykonajmy to samo dla coraz większych zbiorów — po 50, 100, 500, 1000 itd. przypadków i porównajmy ze sobą otrzymane średnie różnic między częstościami a prawdopodobieństwem. Te średnie powinny wypaść coraz mniejsze.

Takie zliczenia i rachunki były niejednokrotnie wykonywane. Dla większych zbiorów były one jednak nieliczne, co nie jest wcale dziwne wobec ogromnej ich uciążliwości. Przytoczę poniżej dane, które udało mi się zebrać. Wszystkie one odnoszą się do prawdopodobieństwa $\frac{1}{2}$.

Zacznę od danych ogłoszonych przez Charliera w jego podręczniku (zob. ust. 4). Ten autor przy pomocy swoich przyjaciół wykonał 10 tysięcy ciągnień z talii kart i notował czarne karty. Materiał został rozbity na zbiory po 10, 50 i 500 ciągnień (tab. 22,1-3).

Tabela 22,1

Czarne karty w zbiorach po 10 ciągnień

Liczby czarnych kart	Ilości zbiorów, w których one wystąpiły
0	3
1	10
2	43
3	116
4	221
5	247
6	202
7	115
8	34
9	9
10	0
Ogółem	1000

Tabela 21,2

Czarne karty w zbiorach po 50 ciągnień

Liczby czarnych kart	Ilości zbiorów, w których one wystąpiły
14	1
15	0
16	2
17	2
18	4
19	8
20	6
21	15
22	13
23	15
24	34
25	14
26	21
27	26
28	14
29	10
30	5
31	5
32	3
33	2
Ogółem	200

Tabela 21,3

Czarne karty w 20 zbiorach po 500 ciągnień

252, 235, 248, 271, 260, 246, 228, 229, 234, 250,
271, 234, 258, 233, 273, 244, 249, 241, 231, 245.

Z danych tabeli 21,3 można utworzyć 10 zbiorów po 1000 ciągnień, dodając po dwie kolejne liczby. Wypadnie: 487, 519, 506, 457, 484, 505, 491, 517, 490, 477. Oprócz tego można utworzyć 2 zbiory po 4000 ciągnień, dodając po 8 liczb. Wypadną w nich liczby czarnych kart 1969 i 1975.

Podobne doświadczenie wykonał W e s t e r g a a r d, ciągnąc kule z worka, zawierającego równe ilości białych i czerwonych kul (cytuję za A n d e r s o n e m). Ciągnień było tak samo 10 tysięcy. Zostały one zgrupowane w zbiory po sto. Liczby białych kul w tych zbiorach są zestawione w tabeli 21,4.

Tabela 21,4

Liczby białych kul w zbiorach po 100 ciągnień.

Liczby białych kul	Ilości zbiorów, w których te liczby wystąpiły
34	1
39	1
40	2
41	2
42	2
43	3
44	3
45	4
46	5
47	6
48	5
49	11
50	9
51	5
52	10
53	4
54	8
55	3
56	5
57	4
58	4
61	1
62	1
63	1
Ogółem	100

Dalej mamy przytoczone na początku czterotysięczne zbiory Buffona, de Morgana i Queteleta.

Przychodzą z kolei zbiory dziesięciotysięczne. Dwa takie doświadczenia ogłosił Jevons w r. 1887 (cytuję za Romanowskim). Było to rzucanie monet po 10240 razy. Jedno z doświadczeń jest obciążone błędem (być może po prostu chochlik drukarski). Drugie dało 5222 razy orła, a więc częstość 0.5100. Do tego dochodzi doświadczenie Charliera z kartami. Dało ono 4933 razy czarną kartę na równe 10 tysięcy ciągnień. Wreszcie doświadczenie Westergaarda dało 5011 razy białą kulę także na 10000 ciągnień.

Bardzo ciekawe dane dla rozpatrywanego zagadnienia ogłosił Marbe[1]) odnośne do wyników gry w ruletkę. Ruletka posiada, jak wiadomo, równe ilości pól czerwonych i czarnych, nadto jedno pole zerowe. Prawdopodobieństwo czerwieni wypada równe $1/2$, tak samo jak w grze w orła i reszkę, jeżeli nie liczyć wystąpień zera. Marbe obliczył dla różnych zbiorów częstości wystąpienia czerwieni dla ruletek w Monte Carlo i w Baden-Baden. Wybieram z tych danych niektóre liczby

Monte Carlo	4253	Częstość czerwieni	0.5079
	10479		0.5050
	29922		0.4991
	58597		0.4984
Baden-Baden	4142		0.5171
	9865		0.5122
	29703		0.5070

Wreszcie Romanowski ogłosił w r. 1912 w „Warszawskich Uniwersitetskich Izwiestjach" wyniki swoich doświadczeń z 80640 rzutami monet. Orzeł wystąpił 39702 razy a więc z częstością 0.4923.

---

[1]) Karl Marbe. Grundfragen der angewandten Wahrscheinlichkeitsrechnung und theoretischen Statistik. — München (1934).

Obliczając średnie wartości różnic między częstościami a prawdopodobieństwem, otrzymujemy dla tych wszystkich danych następujące zestawienie (tab. 21,5).

Tabela 21,5

Wielkości zbiorów w przybliżeniu	Ich liczby	Średnie różnic między częstościami a prawdopodobieństwem
10	1000	0.1107
50	200	0.0561
100	100	0.0409
500	20	0.0227
1000	10	0.0161
4000	7	0.0046
10000	5	0.0070
30000	2	0.0019
58000	1	0.0016
80000	1	0.0077

Widzimy, że na ogół doświadczenie potwierdza prawo wielkich liczb. Dla wielkich zbiorów jednak kwestia pozostaje niewyjaśniona, a to z dwóch powodów. Po pierwsze doświadczenia z takimi zbiorami są nieliczne i przeto niemiarodajne, bo najmniej nawet prawdopodobne zdarzenia czasem występują. Po wtóre w doświadczeniach z wielkimi zbiorami trzeba mieć na uwadze to, że prawo wielkich liczb zachowuje swoją wartość tylko wtedy, jeżeli zdarzenia występują dokładnie w tych samych warunkach. Otóż przy wielkich zbiorach przypadków jest to praktycznie niewykonalne. Jeżeli zdarzenia dotyczą występowania takich czy innych kart w ciągnieniach, sposób mieszania ich powinien być ciągle taki sam, co niekoniecznie ma miejsce. Nadto stan kart powinien być ten sam, a przecież one ścierają się i stają się lepkie przez ciągłe zetknięcie się z dłońmi. Podobnie jest w doświadczeniu z rzucaniem monet — ścierają się one przy uderzeniach o stół itd.

22. Prawdopodobieństwo całkowite i złożone. W poprzednim ustępie omawialiśmy najprostszą a zarazem podstawową formę prawdopodobieństwa, które dotyczyło pojedyńczego zdarzenia. Zagadnienia prawdopodobieństwa mogą oczywiście dotyczyć także dwu i więcej zdarzeń. Zdarzenia te mogą wykluczać się wzajemnie albo przeciwnie występować łącznie. Stosownie do tego będziemy mieli oprócz prawdopodobieństwa prostego — prawdopodobieństwo całkowite i prawdopodobieństwo złożone. Trzeba umieć obliczać te bardziej złożone formy prawdopodobieństwa na podstawie prawdopodobieństwa prostego poszczególnych zdarzeń.

Prawdopodobieństwo całkowite dotyczy zdarzeń wykluczających się wzajemnie. Obliczenie wartości prawdopodobieństwa jest w tym przypadku bardzo proste. Jeżeli z dwóch rodzajów zdarzeń wykluczających się wzajemnie jedno ma prawdopodobieństwo p_1, a drugie p_2, to prawdopodobieństwo całkowite, że zajdzie albo pierwsze, albo drugie zdarzenie, jest równe ich sumie: $p = p_1 + p_2$. Naprzykład, jakie jest prawdopodobieństwo, że z talii kart wyciągniemy albo pika, albo czerwonego asa? Prawdopodobieństwo pierwszego zdarzenia jest $\frac{13}{52}$, drugiego zaś $\frac{2}{52}$. Prawdopodobieństwo, że ciągnąc kartę natrafimy albo na pika, albo na czerwonego asa, jest oczywiście $\frac{13}{52} + \frac{2}{52} = \frac{15}{52}$. Podobnie będzie, jeżeli będzie chodziło o większą ilość różnych a wykluczających się rodzajów zdarzeń.

Z prawidła całkowitego prawdopodobieństwa wynika, że jeżeli weźmiemy wszystkie możliwe w pewnym zakresie rodzaje zdarzeń, to suma ich prawdopodobieństw będzie równa jedności. Istotnie wtedy którekolwiek zdarzenie wystąpi z pewnością a pewność to jest prawdopodobieństwo równe jedności. Stąd wypływa ważny wniosek dotyczący prawdopodobieństwa niewystąpienia zdarzenia. Jeżeli prawdopodobień-

stwo wystąpienia zdarzenia jest p, to prawdopodobieństwo q niewystąpienia go będzie równe $1-p$. Naprzykład można postawić pytanie, jakie jest prawdopodobieństwo niewyciągnięcia pika. Ponieważ prawdopodobieństwo wyciągnięcia takiej karty równa się $\frac{13}{52} = \frac{1}{4}$, prawdopodobieństwo wystąpienia jakiegokolwiek innego koloru jest $1 - \frac{1}{4} = \frac{3}{4}$.

Przechodzimy z kolei do prawdopodobieństwa złożonego. Odnosi się ono do częstości łącznego występowania dwóch lub wielu niezależnych od siebie rodzajów zdarzeń, np. do wyciągnięcia czarnej karty z jednej talii i jednocześnie czerwonego asa z drugiej. Otóż prawdopodobieństwo złożone równa się iloczynowi prawdopodobieństw danych rodzajów zdarzeń.

Jeżeli dajmy na to mamy dwa rodzaje zdarzeń i prawdopodobieństwo jednego z nich jest p_1 a drugiego p_2, to prawdopodobieństwo łącznego wystąpienia tych dwóch rodzajów zdarzeń będzie równe $p = p_1 p_2$.

A więc w przytoczonym przykładzie prawdopodobieństwo wyciągnięcia czarnej karty równa się $\frac{1}{2}$, prawdopodobieństwo zaś wyciągnięcia czerwonego asa jest $\frac{1}{26}$. Przeto prawdopodobieństwo łącznego wystąpienia tych zdarzeń równa się $\frac{1}{2} \times \frac{1}{26} = \frac{1}{52}$. W istocie, ponieważ każda karta jednej talii może spotkać się z każdą kartą drugiej, ilość różnych jednakowo możliwych ewentualności jest 52×52. Wśród nich każda z 26 czarnych kart może spotkać się z każdym z 2 czerwonych asów. Zatem liczba różnych spotkań tego rodzaju jest 26×2. W ten sposób prawdopodobieństwo spotkania się czarnej karty z czerwonym asem wypada równe $\frac{26 \times 2}{52 \times 52} = \frac{1}{52}$. Tak samo można rozumować w każdym innym przypadku.

23. Prawdopodobieństwo powtarzania się zdarzeń. Szereg dwumianowy. Podane w poprzednim ustępie twierdzenie o prawdopodobieństwie złożonym może być stosowane nie tylko do łącznego występowania różnych rodzajów zdarzeń, lecz także do łącznego występowania tego samego rodzaju zdarzeń, a więc może być użyte do obliczenia prawdopodobieństwa powtórzeń tego samego zdarzenia. To zagadnienie było opracowane na przełomie XVII i XVIII wieku przez B e r n o u l l i'e g o, o którym była już mowa w ust. 21.

Rozpatrzmy z początku najprostszą formę zagadnienia, kiedy zdarzenie występuje najwyżej dwukrotnie. Tym zdarzeniem niech będzie występowanie czarnej karty w ciągnieniach z talii. A więc ciągnijmy z dwóch talii po jednej karcie albo, co na jedno wyjdzie, ciągnijmy z tej samej talii powtórnie po włożeniu do niej z powrotem wyciągniętej poprzednio karty i po przetasowaniu talii. Będziemy mieli cztery równie możliwe ewentualności: 1) pierwsze i drugie ciągnienie dadzą czerwone karty, 2) pierwsze da czerwoną, a drugie czarną, 3) pierwsze czarną a drugie czerwoną i wreszcie 4) oba ciągnienia dadzą czarne karty. Oznaczymy zjawienie się czarnej karty znakiem +, czerwonej znakiem —. Będziemy wtedy mogli zestawić omawiane ewentualności w formie następującej:

1) — —
2) — +
3) + —
4) + +

Są tu zestawione we wszelki możliwy sposób dwa różne rodzaje zdarzeń. Podobne zestawienia noszą nazwę p r z e m i a n (wariacyj). Zestawienie zbiorów tak mało licznych jest proste. Natomiast dla większych zbiorów zadanie staje się o wiele trudniejsze, zwłaszcza jeżeli są rozpatrywane nie dwie tylko jak tu ewentualności, lecz większa ich ilość. Istnieje nawet osobna gałęź matematyki, k o m b i n a t o r y k i[1]), która się tym zajmuje.

[1]) Najbardziej wyczerpującym źródłem w tej dziedzinie jest książka E. N e t t o. Lehrbuch der Combinatorik (Lipsk, 2 wyd. 1927).

Widzimy odrazu, że przemiana 1 nie da żadnej czarnej karty, przemiany 2 i 3 dadzą jedną czarną kartę, wreszcie przemiana 4 da dwie karty czarne. Prawdopodobieństwo każdej przemiany jest według prawidła złożonego prawdopodobieństwa $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Składając prawdopodobieństwa przemian 2 i 3, otrzymamy dla jednorazowego wystąpienia czarnej karty prawdopodobieństwo podwójne. Ostatecznie wypadnie następujący szereg prawdopodobieństw:

Liczby czarnych kart w zbiorach	0	1	2
Ich prawdopodobieństwa	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Suma tego szeregu wypada równa jedności, co jest naturalne, bo są tu uwzględnione wszystkie możliwe ewentualności, jakie mogą wystąpić w dwójkowych zbiorach ciągnień.

Ciągnijmy następnie karty trzykrotnie — będziemy mieli trójkowe zbiory przypadków. Liczba jednakowo możliwych zestawień będzie teraz dwa razy większa niż poprzednio. Istotnie każdej przemianie dwójkowych zbiorów będą tu odpowiadały dwie, bo w trzecim ciągnieniu może wyjść zarówno czerwona jak czarna karta. Oznaczając literami *a* i *b* ewentualności trzeciego ciągnienia, możemy zestawić następujący zespół przemian:

1 *a*)	—	—	—
1 *b*)	—	—	+
2 *a*)	—	+	—
2 *b*)	—	+	+
3 *a*)	+	—	—
3 *b*)	+	—	+
4 *a*)	+	+	—
4 *b*)	+	+	+

Teraz przemiana 1*a* nie da żadnej czarnej karty, przemiany 1*b*, 2*a* i 3*a* dadzą jedną czarną kartę, 2*b*, 3*b* i 4*a* — dadzą dwie i wreszcie przemiana 4*b* — trzy czarne karty. Prawdopodobieństwo każdej przemiany jest $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

Składając prawdopodobieństwa, otrzymamy szereg prawdopodobieństw dla trójkowych zbiorów:

Liczby czarnych kart w zbiorach:	0	1	2	3
Ich prawdopodobieństwa:	$\frac{1}{8}$	$3\times\frac{1}{8}$	$3\times\frac{1}{8}$	$\frac{1}{8}$

Suma wyrazów szeregu znowu równa jest jedności.

Zastosujemy z kolei takie rozumowanie do czwórkowych zbiorów ciągnień. Znowu każdej przemianie poprzedniego zadania będą odpowiadały dwie przemiany *a* i *b* nowego. Otrzymamy 16 przemian:

1 *aa)*	—	—	—	—
1 *ab)*	—	—	—	+
1 *ba)*	—	—	+	—
1 *bb)*	—	—	+	+
2 *aa)*	—	+	—	—
2 *ab)*	—	+	—	+
2 *ba)*	—	+	+	—
2 *bb)*	—	+	+	+
3 *aa)*	+	—	—	—
3 *ab)*	+	—	—	+
3 *ba)*	+	—	+	—
3 *bb)*	+	—	+	+
4 *aa)*	+	+	—	—
4 *ab)*	+	+	—	+
4 *ba)*	+	+	+	—
4 *bb)*	+	+	+	+

Prawdopodobieństwo każdej przemiany będzie $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{16}$. Zliczając przemiany o jednakowej liczbie czarnych kart, ułożymy szereg prawdopodobieństw:

Liczby czarnych kart w zbiorach:	0	1	2	3	4
Ich prawdopodobieństwa:	$\frac{1}{16}$	$4\times\frac{1}{16}$	$6\times\frac{1}{16}$	$4\times\frac{1}{16}$	$\frac{1}{16}$

Rozumując w podobny sposób dla zbiorów po 5 kart, otrzymamy szereg:

Liczby czarnych kart w zbiorach:	0	1	2	3	4	5
Ich prawdopodobieństwa:	$\frac{1}{32}$	$3\times\frac{1}{32}$	$10\times\frac{1}{32}$	$10\times\frac{1}{32}$	$5\times\frac{1}{32}$	$\frac{1}{32}$

Tak samo można rozumować dalej przy coraz większej ilości ciągnień. Zawsze przy tym suma prawdopodobieństw będzie równa jedności.

Porównując ze sobą otrzymane szeregi prawdopodobieństw, widzimy, że ich współczynniki są takie same, jak w rozwinięciu dwumianu Newtona takiego samego stopnia jak liczba przypadków w danych zbiorach. Oznaczając prawdopodobieństwo wyjścia czarnej karty przez p, prawdopodobieństwo zaś wyjścia czerwonej, to znaczy nie wyjścia czarnej, przez q, możemy odnośne rozwinięcia napisać w formie:

$$(q + p)^2 = q^2 + 2qp + p^2$$
$$(q + p)^3 = q^3 + 3q^2p + 3qp^2 + p^3$$
$$(q + p)^4 = q^4 + 4q^3p + 6q^2p^2 + 4qp^3 + p^4$$
$$(q + p)^5 = q^5 + 5q^4p + 10q^3p^2 + 10q^2p^3 + 5q^4 + p^5$$

Nietrudno jest stwierdzić, że te wielomiany nietylko mają takie same współczynniki jak rozpatrywane szeregi prawdopodobieństw, lecz dają wprost wyraz po wyrazie odnośne prawdopodobieństwa, jeżeli wstawimy w miejce p i q ich liczbowe wartości, które tu są obie równe ½. Stąd te szeregi noszą nazwę dwumianowych. Nazywa się je także szeregami Bernoulli'ego.

To jeszcze nie wszystko. Rozwinięcie dwumianu Newtona daje prawdopodobieństwa powtórzeń zdarzenia także wtedy, kiedy prawdopodobieństwo wystąpienia go nie jest równe prawdopodobieństwu niezjawienia się. Rozpatrzmy naprzykład prawdopodobieństwo wyciągnięcia nie byle jakiej czarnej karty, lecz specjalnie pika. Prawdopodobieństwo zdarzenia będzie teraz ¼ a nie wystąpienia go $\frac{3}{4}$.

Możemy tu wykorzystać te same zestawienia przemian, jakie mieliśmy poprzednio, ale ich prawdopodobieństwa nie będą jednakowe. I tak oznaczając literą p z odpowiednim znaczkiem ich prawdopodobieństwa, będziemy mieli dla dwójkowych zbiorów: $p_1 = \frac{3}{4} \times \frac{3}{4} = \left(\frac{3}{4}\right)^2$, $p_2 = p_3 = \frac{3}{4} \times \frac{1}{4}$, $p_4 = \frac{1}{4} \times \frac{1}{4} = \left(\frac{1}{4}\right)^2$.

Wypadnie wtedy szereg prawdopodobieństw według dwumianu drugiego stopnia $(q + p)^2$, jeżeli w nim wstawimy $q = \frac{3}{4}$, $p = \frac{1}{4}$, a mianowicie:

Liczby pików w zbiorach:	0	1	2
Ich prawdopodobieństwa:	$\left(\frac{3}{4}\right)^2$	$2 \times \frac{3}{4} \times \frac{1}{4}$	$\left(\frac{1}{2}\right)^2$

W trójkowych zbiorach otrzymamy dla różnych występujących w nich przemian prawdopodobieństwa, które można zestawić w sposób następujący, jeżeli się oznaczy je po prostu znaczkami przemian w kwadratowych nawiasach:

$$[1a] = \left(\frac{3}{4}\right)^3$$

$$[1b] = [2a] = [3a] = \left(\frac{3}{4}\right)^2 \times \frac{1}{4}$$

$$[2b] = [3b] = [4b] = \frac{3}{4}\left(\frac{1}{4}\right)^2$$

$$[4b] = \left(\frac{1}{4}\right)^3$$

Wypadnie zatem szereg prawdopodobieństwa według dwumianu trzeciego stopnia $\left(\frac{3}{4} + \frac{1}{4}\right)^3$.

Liczby pików w zbiorach:	0	1	2	3	4
Ich prawdopodobieństwa:	$\left(\frac{3}{4}\right)^3$	$3 \times \left(\frac{3}{4}\right)^2 \times \frac{1}{4}$	$3 \times \frac{3}{4} \times \left(\frac{1}{4}\right)^2$	$\left(\frac{1}{4}\right)^3$	

Tak samo można przerobić czwórkowe zbiory. Będziemy mieli następujące prawdopodobieństwa różnych przemian:

$$[1aa] = \left(\frac{3}{4}\right)^4$$

$$[1ab] = [1ba] = [2aa] = [3aa] = \left(\frac{3}{4}\right)^3 \times \frac{1}{4}$$

$$[1bb] = [2ab] = [2ba] = [3ab] = [4aa] = \left(\frac{3}{4}\right)^2 \times \left(\frac{1}{4}\right)^2$$

$$[2bb] = [3bb] = [4ab] = [4ba] = \frac{3}{4} \times \left(\frac{1}{4}\right)^3$$

$$[4bb] = \left(\frac{1}{4}\right)^4$$

Szereg prawdopodobieństw wypadnie zgodnie z dwumianem $\left(\frac{3}{4} + \frac{1}{4}\right)^4$:

Liczby pików w zbiorach:	0	1	2	3	4
Ich prawdopodobieństwa	$\left(\frac{3}{4}\right)^4$	$4\times\left(\frac{3}{4}\right)^3\times\frac{1}{4}$	$6\times\left(\frac{3}{4}\right)^2\times\left(\frac{1}{4}\right)^2$	$4\times\frac{3}{4}\times\left(\frac{1}{4}\right)^3$	$\left(\frac{1}{4}\right)^4$

Tak samo będzie dla zbiorów jeszcze większych.

Porównując ze sobą szeregi dwumianowe widzimy, że prawdopodobieństwo niewystąpienia zdarzenia jest równe pierwszemu wyrazowi szeregu, prawdopodobieństwo jednokrotnego wystąpienia równe drugiemu wyrazowi, dwukrotnego — trzeciemu, trzykrotnego — czwartemu itd. Dochodzimy w ten sposób do ogólnego prawidła:

Jeżeli prawdopodobieństwo wystąpienia jakiegoś zdarzenia jest p a prawdopodobieństwo niewystąpienia go $q = 1 - p$, to prawdopodobieństwo, że zdarzenie wystąpi k razy w zbio-

rach po n przypadków, jest równe $(k+1)$-nemu wyrazowi dwumianu $(q+p)^n$. Takie prawdopodobieństwo wyraża się wzorem:

$$\frac{n\,(n-1)\,.\,.\,.\,(n-k+1)}{1.\,2.\,3.\,.\,.\,.\,k}\quad q^{n-k}p^k$$

Ten wzór można napisać w formie bardziej przejrzystej, jeżeli użyje się tzw. silni, iloczynów kolejnych liczb całkowitych od jedynki począwszy. Silnie oznacza się ostatnią liczbą z wykrzyknikiem, np. 1. 2. 3. 4. 5. przez 5! Rozpatrywany wzór będzie wtedy:

$$\frac{n!}{k!\,(n-k)!}\quad q^{n-k}p^k$$

Przy stosowaniu jego trzeba mieć na uwadze, że silnia 0! jest równa jedności.

Do obliczenia wyrazów szeregu dwumianowego wypada ciągle obliczać współczynniki dwumianu Newtona. Jest to bardzo kłopotliwe przy wysokich potęgach. Dla ułatwienia można przytem korzystać z tzw. trójkątu Pascala. Układa się go w sposób następujący. Piszemy w pierwszym wierszu dwie jedynki. W drugim wierszu piszemy między nimi ich sumę i dopisujemy z obu stron nowe jedynki:

1 1
1 2 1

Dodajemy następnie stojące obok siebie liczby drugiego wiersza i wpisujemy sumy w trzecim wierszu pomiędzy odnośnymi liczbami drugiego. Oprócz tego dopisujemy znowu jedynki po obu stronach:

1 1
1 2 1
1 3 3 1

Postępując tak samo dalej, ułożymy trójkąt Pascala.

który naturalnie można przedłużyć do nieskończoności. Poniżej podaję go w formie doprowadzonej do 10-go stopnia:

1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 79 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1

By otrzymać współczynniki dwumianu k-ego stopnia trzeba wziąć wiersz numer k.

Suma współczynników dla dwumianu k-ego stopnia równa się 2^k, np. dla dwumianu 10 sotpnia wynosi 1024. Istotnie mamy $(1+1)^k = 2^k$.

Przy bardzo wysokich potęgach nawet użycie trójkąta Pascala jest zbyt mozolne. Można wtedy sobie radzić, ujmując spółczynniki za pomocą silni a więc w formie $\frac{n!}{k!\,(n-k)!}$. Jest to wygodne dzięki wzorowi angielskiego matematyka Stirlinga, który daje możność przybliżonego wprawdzie ale wygodnego obliczenia silni. Wzór ten jest następujący:

$$n! = 1.2.3.\ldots n = \sqrt{2\pi n} \times n^n \times e^{-n}$$

W nim e jest przestępną liczbą, którą można obliczyć za pomocą nieskończonej sumy:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots.$$

Jej przybliżona wartość wynosi: 2.718251828459045...

Porównując ze sobą różne szeregi dwumianowe, łatwo jest stwierdzić, że jeżeli prawdopodobieństwo wystąpienia zdarzenia jest równe połowie a więc jeżeli $q = p$, to szereg będzie symetryczny, w przeciwnym razie — skośny. I tak dla czwórkowych

zbiorów przy $p = {}^1/_2$ wypadł nam szereg (wyrażony w szesnastych częścach jedności): 1, 4, 6, 4, 1. Natomiast przy $p = {}^1/_4$ otrzymaliśmy (w 256-ych częściach jedności) : 81, 108, 54, 12, 1. Skośność w drugim przypadku wypadła bardzo silna, będzie ona jeszcze silniejsza, jeżeli p i q będą się jeszcze bardziej różniły. Szereg przy dostatecznie dużej różnicy między tymi prawdopodobieństwami może stać się nawet jednobocznym. I tak przy $p = {}^1/_{10}$, $q = {}^9/_{10}$ wypadnie on (w dziesięciotysięcznych częściach jedności) dla czwórkowych zbiorów: 6561, 2916, 486, 36, 1.

Zwiększenie zbiorów wyrównuje skośność w znacznym stopniu nawet przy dużej różnicy między p i q. I tak Yule obliczył następujące prawdopodobieństwa (w dziesięciotysięcznych częściach jedności) dla zbiorów złożonych ze 100 przypadków przy $p = {}^1/_{10}$ (tab. 23,1). W tej tabeli wartości mniejsze od połowy dziesięciotysięcznej są pominięte. Tak małe prawdopodobieństwa wypadają dla wysokich liczb powtórzeń, mianowicie dla 24 do 100. Równie małe jest prawdopodobieństwo nie wystąpienia zdarzenia.

Tabela 23,1

Prawdopodobieństwa różnych powtórzeń zdarzenia o prawdopodobieństwie $\frac{1}{10}$ w zbiorach złożonych ze 100 przypadków (dziesięciotysięczne części jedności).

Liczby powtórzeń	Ich prawdopodobieństwa
0	—
1	3
2	16
3	59
4	159
5	339
6	596
7	889

Liczby powtórzeń	Ich prawdopodobieństwa
8	1148
9	1304
10	1319
11	1199
12	988
13	743
14	513
15	327
16	193
17	106
18	54
19	26
20	12
21	5
22	2
23	1
24	—

Ponieważ szereg dwumianowy obejmuje prawdopodobieństwa wszystkich możliwych zestawień zdarzeń, suma jego wyrazów jest zawsze równa jedności. Daje to możność sprawdzania obliczeń.

Użycie dwumianu Newtona do obliczania prawdopodobieństw można rozszerzyć na zagadnienia, w których chodzi o powtarzanie się nie jednego zdarzenia, lecz dwóch i więcej. Można naprzykład postawić sobie zadanie obliczyć prawdopodobieństwa występowania czarnych kart i kierów. Nie będę omawiał szczegółowo takich zagadnień, gdyż jest to dosyć zawiłe. Podam tylko gotowy wzór.

Niech będą p_1, p_2 ... p prawdopodobieństwa występowania s różnych rodzajów zdarzeń, przy czym suma tych prawdopodobieństw jest równa jedności. Wówczas prawdopodobieństwo wystąpienia a_1 razy pierwszego rodzaju

zdarzeń a jednocześnie a_2 razy zdarzeń drugiego rodzaju itd. i a_s razy ostatniego rodzaju w zbiorach po n przypadków, gdzie $n = a_1 + a_2 + \ldots + a_s$, wyrazi się wzorem:

$$\frac{n!}{a_1!\, a_2! \ldots a_s!}\; p_1^{a_1}\, p_2^{a_2} \cdots p_s^{a_s}$$

Jest to ogólny wzór wyrazu w wielomianie $(p_1 + p_2 + \ldots + p_s)^n$

Dla wyjaśnienia, jak się korzysta z powyższego wzoru, obliczmy jakie jest prawdopodobieństwo, że w zbiorach ciągnień po 5 kart otrzyma się trzy czarne karty i jednego kera. Prawdopodobieństwo czarnej karty jest $1/2$, jest to p_1. Prawdopodobieństwo kera jest $1/4$, mamy zatem $p_2 = 1/4$. Ponieważ suma tych prawdopodobieństw jest mniejsza od jedności, trzeba jeszcze wprowadzić trzecie — prawdopodobieństwo wystąpienia kart nieczarnych i niekerów $p_3 = 1/4$. Odnośne liczby powtórzeń będą: $a_1 = 3$, $a_2 = 1$, $a_3 = 1$. Prawdopodobieństwo tych powtórzeń wypadnie:

$$\frac{5!}{3!\; 1!\; 1!} \left(\frac{1}{2}\right) \times \frac{1}{4} \times \frac{1}{4} = \frac{28}{128}$$

Szereg dwumianowy ma podstawowe znaczenie dla nauki o prawdopodobieństwie i statystyki. Nie dość na tym, że sam przez się odgrywa poważną rolę, ale nadto wywodzą się z niego dwie doniosłe koncepcje — normalna funkcja prawdopodobieństwa zwana inaczej funkcją de Moivre'a i szereg Poissona. W dalszych ustępach zajmiemy się tymi koncepcjami.

24. Charakterystyki szeregu dwumianowego. Szeregi rozdzielcze wielu materiałów statystycznych są podobne do szeregu dwumianowego. W tym szeregu liczby powtórzeń grają rolę zmiennej ewentualnej a ich prawdopodobieństwa rolę częstości. Analogia jest bardzo ścisła, bo prawdopodobieństwo, jak widzieliśmy w ust. 21, jest graniczną wartością częstości przy zwiększaniu liczby przypadków, jeżeli częstość jest wzięta

w formie częstości względnej. Wobec tego jest konieczne ustalenie dla szeregu dwumianowego takich samych charakterystyk, jak dla szeregów rozdzielczych.

Rozpoczniemy od średniej arytmetycznej. Zastosujemy tu naturalnie metodę wartości wyjściowej, przyjmując tę wartość równą zeru. Użyjemy znanego nam schematu rachunkowego (tabl. 24,1).

Tabela 24,1

x (liczby powtórzeń)	f ich prawdopodobieństwa	xf
0	q^n	0
1	$nq^{n-1}p$	$nq^{n-1}p$
2	$\frac{n(n-1)}{1.2} q^{n-2} p^2$	$n(n-1) q^{n-2} p^2$
3	$\frac{n(n-1)(n-2)}{1.2.3.} q^{n-3} p^3$	$\frac{n(n-1)(n-2)}{1.2} q^{n-3} p^3$

Sumując iloczyny xf i wynosząc za nawias wspólny czynnik np, otrzymamy jako sumę odchyleń od wartości wyjściowej wyrażenie:

$$\Sigma xf = np\,[q^{n-1} +](n-1)\, q^{n-2} p + \frac{(n-1)(n-2)}{1.2} q^{n-3} p^2 + \ldots]$$

Wielomian w nawiasie jest niczym innym jak dwumianem Newtona $(q+p)^{n-1}$. Ponieważ suma wyrazów takiego dwumianu równa się jedności, wypadnie $\Sigma\, xf = np$. Dla otrzyotrzymania pomocniczego momentu μ_1 trzeba sumę odchyleń podzielić przez sumę f, a ponieważ jest to suma wyrazów dwumianu $(q+p)^{n-1}$, która jest równa jedności, dzielenie jest zbyteczne i otrzymujemy odrazu $\mu_1 = np$. Taką samą wartość będzie miała średnia arytmetyczna, gdyż wartość wyjściowa jest równa zeru.

Przychodzimy teraz do momentów. Pomijamy częściowe momenty pierwszego stopnia, gdyż do nich obliczenia trzeba oddzielić minus warianty od plus wariantów. Jest to łatwe w poszczególnych przypadkach, natomiast w formie ogólnych wzorów jest niemożliwe. Przechodzimy więc do momentów wyższych rzędów. I tak dla momentu drugiego stopnia będziemy mieli, przyjmując znowu jako wartość wyjściową, zero (tab. 24,2). Sumując iloczyny fx^2 i wyciągając za nawias wspólny czynnik np, otrzymamy:

$$\Sigma fx^2 = np\,[q^{n-1} + 2(n-1)\,q^{n-2}p + 3\,\frac{(n-1)(n-2)}{1.2}\,q^{n-3}p^2 + $$
$$+ 4\,\frac{(n-1)(n-2)(n-3)}{1.2.3.}\,q^{n-4}p^3 \;\ldots]$$

Tabela 24,2

x	f	x^2	fx^2
0	q^n	0	0
1	$nq^{n-1}p$	1	$nq^{n-1}p$
2	$\frac{n(n-1)}{1.2}\,q^{n-2}p^2$	4	$2n(n-1)\,q^{n-2}p^2$
3	$\frac{n(n-1)(n-2)}{1.2.3}\,q^{n-3}p^3$	9	$3\frac{n(n-1)(n-2)}{1.2}\,q^{n-3}p^3$
4	$\frac{n(n-1)(n-2)(n-3)}{1.2.3.4}\,q^{n-4}p^4$	16	$4\frac{n(n-1)(n-2)(n-3)}{1.2.3}\,q^{n-4}p^4$

Wielomian zawarty w nawiasach przekształcamy w ten sposób, by z niego wydzielić dwumian $(q+p)^{n-1}$:

$$\Sigma fx^2 = np\,[q^{n-1} + (n-1)\,q^{n-2}p + \frac{(n-1)(n-2)}{1.2}\,q^{n-3}p^2 +$$
$$+ \frac{(n-1)(n-2)(n-3)}{1.2.3}\,q^{n-4}p^3 + \ldots . + (n-1)\,q^{n-2}p +$$
$$+ 2\,\frac{(n-1)(n-2)}{1.2}\,q^{n-3}p^2 + 3\,\frac{(n-1)(n-2)(n-3)}{1.2.3}\,q^{n-4}p^3 + \ldots]$$

Mamy teraz w nawiasach dwa wielomiany. Pierwszy z nich jako suma wyrazów szeregu dwumianowego jest równy jedności. Co do drugiego, porównajmy go z wyprowadzoną poprzednio sumą odchyleń od zera wyrazów szeregu $(q+p)^n$. Widzimy, że ten drugi wielomian da się wyprowadzić z wyrażenia dla wspomnianej sumy, jeżeli n zastąpi się przez $n-1$. Jest to zatem średnia arytmetyczna szeregu $(q+p)^{n-1}$ a więc ma wartość $(n-1)p$. W ten sposób dla sumy iloczynów fx^2 wypadnie wyrażenie $np\,[+(n-1)p]$, czyli $np+n^2p^2-np^2$. Otrzymaliśmy w ten sposób pomocniczy moment μ_2, gdyż dzielenie przez $\Sigma\, f=1$ jest zbyteczne. Ostatecznie dla momentu drugiego stopnia otrzymamy:

$$m_2=\mu_2-\mu_1^2=np+n^2p^2-np^2-n^2p^2=np-np^2=np(1-p)=npq$$

Stosując tę samą metodę do momentów trzeciego i czwartego stopnia, otrzymamy:

$$m_3=npq\,(p-q)$$

$$m_4=3n^2p^2q^2+npq\,(1-6pq)$$

Wyprowadzenie tych wzorów jest zbyt zawiłe, by je tu przytaczać.

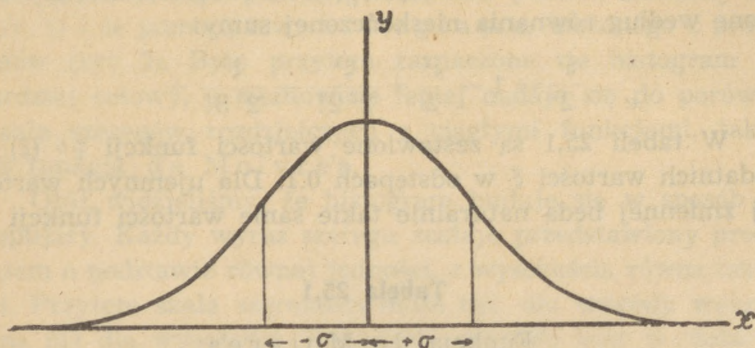
25. Normalna funkcja prawdopodobieństwa. Widzieliśmy w ust. 23, że w szeregu dwumianowym skośność powodowana przez różnicę między prawdopodobieństwami p i q wyrównuje się przy należytym zwiększeniu n. Mało tego, można dowieść, że przy nieograniczonym zwiększaniu n szereg dwumianowy zbliża się do pewnej stałej postaci, określonej funkcją

$$y=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{x^2}{2\sigma^2}}$$

W tym równaniu e jest znaną już nam liczbą przestępną a wielkość σ jest stałą wielkością, która może mieć różne war-

tości. Ta funkcja nosi nazwę normalnej funkcji prawdopodobieństwa albo funkcji de Moivre'a. Zwykle przypisuje się ją Laplace'owi albo Gaussowi. Jest to niesłuszne. Jak to stwierdził Pearson, de Moivre zajmował się nią o wiele wcześniej — od r. 1733 w dziele „Miscellanea analytica".

Dla należytego wyjaśnienia związku między szeregiem dwumianowym a funkcją de Moivrea, trzeba użyć przedstawienia graficznego (ryc. 3).



Ryc. 3

Widzimy na takim wykresie, że krzywa funkcji wznosi się najwyżej dla $x = 0$. Od tego punktu opada ona symetrycznie w obie strony, idąc nieskończoność. Z początku obie gałęzie są wklęsłe względem osi odciętych, następnie począwszy od $x = + \sigma$ dla prawej gałęzi i od $x = - \sigma$ dla lewej stają się wypukłe. Jedna i druga stopniowo i bezgranicznie, asymptotycznie jak się to mówi, zbliżają się do osi odciętych. Kształt krzywej przypomina w pewnym stopniu klosz albo według dowcipnego francuskiego określenia kapelusz żandarma. Ten „kapelusz" jest tym bardziej wysmukły, im mniejsza jest wielkość σ, z jej zwiększeniem przypłaszcza się.

Wobec wielkiego znaczenia funkcji de Moivre'a obliczono z wielką dokładnością jej wartości, biorąc je w formie

$$\eta = \varphi_0(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$$

w której stała σ jest wyłączona. Taką formę przybiera równanie funkcji przez zmianę zmiennych: $x = \sigma \xi$, $y = \frac{\eta}{\sigma}$. Zmienne ξ i η noszą nazwę normalnych. Za pomocą przytoczonych wzorów można przy danym σ obliczyć wartość y dla każdego x według wartości funkcji $\varphi_0(\xi)$. Te wartości zostały obliczone według równania nieskończonej sumy:

$$e^{-\frac{\xi^2}{2}} = 1 - \frac{\xi^2}{2} + \frac{\xi^4}{2^2 2!} - \frac{\xi^6}{2^3 3!} + \dots$$

W tabeli 25,1 są zestawione wartości funkcji $\varphi_0(\xi)$ dla dodatnich wartości ξ w odstępach 0.1. Dla ujemnych wartości tej zmiennej będą naturalnie takie same wartości funkcji jak

Tabela 25,1

Funkcja de Moivre'a

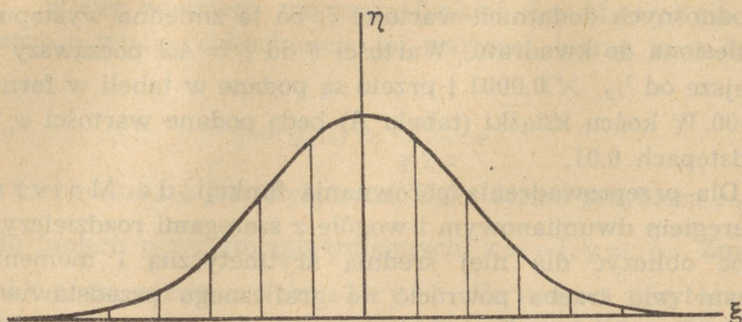
ξ	η	ξ	η	ξ	η	ξ	η
0.0	0.3989	1.3	0.1714	2.6	0.0136	3.8	0.0003
0.1	3970	1.4	1497	2.7	0104	3.9	0002
0.2	3910	1.5	1295	2.8	0079	4.0	0001
0.3	3814	1.6	1190	2.9	0060	4.1	0001
0.4	3683	1.7	0940	3.0	0044	4.2	0001
0.5	3521	1.8	0790	3.1	0033	4.3	0000
0.6	3332	1.9	0656	3.2	0024	4.4	0000
0.7	3123	2.0	0540	3.3	0017	4.5	0000
0.8	2897	2.1	0440	3.4	0012	4.6	0000
0.9	2661	2.2	0355	3.5	0009	4.7	0000
1.0	2420	2.3	0283	3.6	0006	4.8	0000
1.1	2179	2.4	0224	3.7	0004	4.9	0000
1.2	1942	2.5	0175				

dla odnośnych dodatnich wartości ξ, bo ta zmienna występuje podniesiona do kwadratu. Wartości η od $\xi = 4.3$ począwszy są mniejsze od $^1/_2 \times 0.0001$ i przeto są podane w tabeli w formie 0.0000. W końcu książki (tabela *A*) będą podane wartości $\varphi_0(\xi)$ w odstępach 0.01.

Dla przeprowadzenia porównania funkcji de Moivre'a z szeregiem dwumianowym i wogóle z szeregami rozdzielczymi trzeba obliczyć dla niej średnią arytmetyczną i momenty. W tym celu trzeba powrócić do graficznego przedstawienia szeregów rozdzielczych, omówionego w ust. 5. Były tam podane dwa sposoby tego przedstawienia: za pomocą łamanej linii (ryc. 1) i za pomocą tzw. histogramu, złożonego z prostokątów (ryc. 2). Było przytem zaznaczone, że histogram jest bardziej celowy, a mianowicie lepiej nadaje się do porównywania szeregów rozdzielczych z ciągłymi funkcjami, takimi jak funkcja de Moivre'a.

Otóż widzieliśmy, że histogram buduje się w sposób następujący. Każdy wyraz szeregu zostaje przedstawiony prostokątem o podstawie równej jedności, z wysokością równą częstości. Przytem skala szerokości może być dla wygody wybrana inna niż dla wysokości. Ponieważ szerokość klas w materiale niegrupowanym jest równa jedności, pole każdego prostokąta w histogramie równa się odnośnej częstości a suma tych pól równa się liczbie wariantów. To samo będzie z materiałem zgrupowanym, o ile przyjmie się szerokość klas za jednostkę do pomiaru wartości wariantów, co prowadzi do użycia „numerów" do oznaczenia wyników takich pomiarów.

Zastosujmy ten sam sposób rozumowania do krzywej przedstawiającej funkcję de Moivre'a. Dzielimy figurę ograniczoną przez krzywą i oś odciętych na wycinki za pomocą szeregu rzędnych (ryc. 4). Te wycinki o kształcie zbliżonym do trapezów będą odpowiednikami prostokątów histogramu. Wszystkie ich boki, z wyjątkiem górnego, będą prostoliniowe.
Podstawa każdego wycinka będzie odpowiednikiem przedziału klasowego w szeregach rozdzielczych, pole jego — odpowiednikiem odnośnej częstości.



Ryc. 4

Pola podanych powyżej wycinków krzywej funkcji de Moivre'a, ograniczonych przez dwie sąsiednie rzędne, można obliczyć zgrubsza, przyjmując wycinek za trapez. Jeżeli odnośne rzędne mają wartości ξ_2 i ξ_1 ($\xi_2 > \xi_1$), to pole wycinka będzie w przybliżeniu równe

$$\frac{[\varphi_0(\xi_2) + \varphi_0(\xi_1)]\,(\xi_2 - \xi_1)}{2}$$

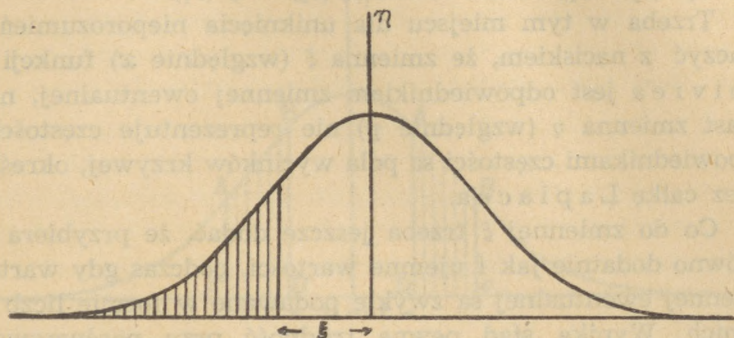
Naprzykład jeżeli $\xi_2 = 1.1$, $\xi_1 = 1.0$, otrzymamy na podstawie tabeli 25,1 wielkość pola wycinka równą

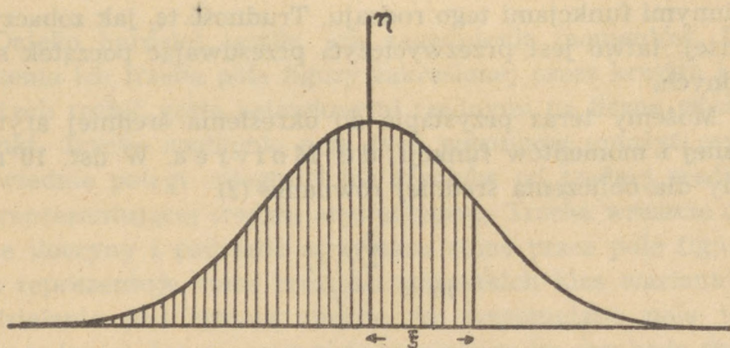
$$\frac{[0.2179 + 0.2420] \times 0.1}{2} = 0.0230$$

W tym wypadku wypada wartość dokładna do czwartego znaku dziesiętnego. Niezawsze tak jest. Ogólnie wynik będzie tym dokładniejszy, im rzędne są bliższe a więc wycinek węższy. Dokładne wartości pól dla wycinków dowolnej szerokości można otrzymać tylko za pomocą rachunku całkowego, obliczając tzw. całką Laplace'a

$$Q(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{\frac{-\xi^2}{2}} . d\xi$$

Daje ona wielkość pola figury ograniczonej przez oś odciętych, l e w ą część krzywej i rzędną dla danej wartości ξ (ryc. 5). Dla $\xi = -\infty$ całka





Ryc. 5

L a p l a c e'a równa się zeru, dla $\xi = 0$ przybiera wartość $1/2$, a przy nieograniczonym zwiększeniu ξ zbliża się do jedności. Wartości tej całki oblicza się za pomocą nieskończonej sumy:

$$Q(\xi) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(\xi - \frac{\xi^3}{6} + \frac{\xi^5}{40} - \frac{\xi^7}{336} + \ldots \right)$$

z ogólnym wyrazem $(-1)^k \dfrac{\xi^{2k+1}}{2^k!\,(2k+1)}$

Za pomocą tabeli *B* podaną w końcu książki pola wycinków dadzą się łatwo i dokładnie obliczyć, odejmując odnośne wartości funkcji $Q(\xi)$. Pole wycinka ograniczonego przez wartości ξ_2 i ξ_1 $(\xi_2 > \xi_1)$ będzie równe $Q(\xi_2) - Q(\xi_1)$.

Trzeba w tym miejscu dla uniknięcia nieporozumień zaznaczyć z naciskiem, że zmienna ξ (względnie x) funkcji de Moivre'a jest odpowiednikiem zmiennej ewentualnej, natomiast zmienna η (względnie y) nie reprezentuje częstości — odpowiednikami częstości są pola wycinków krzywej, określone przez całkę Laplace'a.

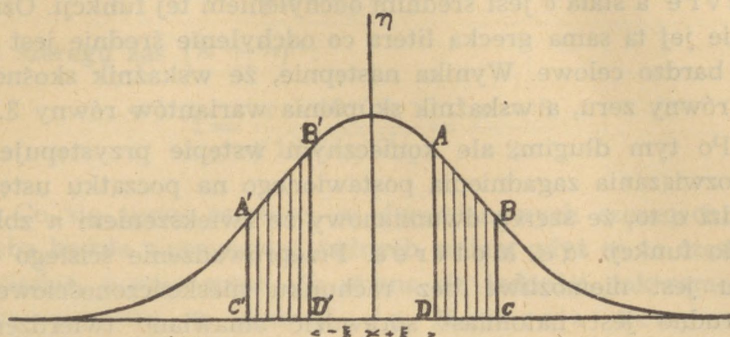
Co do zmiennej ξ trzeba jeszcze dodać, że przybiera ona zarówno dodatnie jak i ujemne wartości, podczas gdy wartości zmiennej ewentualnej są zwykle podawane w formie liczb dodatnich. Wynika stąd pewna trudność przy porównywaniu szeregów rozdzielczych z funkcją de Moivre'a, jak zresztą i z innymi funkcjami tego rodzaju. Trudność tę, jak zobaczymy poniżej, łatwo jest przezwyciężyć, przesuwając początek spółrzędnych.

Możemy teraz przystąpić do określenia średniej arytmetycznej i momentów funkcji de Moivre'a. W ust. 10 mieliśmy dla obliczenia średniej równanie (2)

$$m = \frac{\sum_{1}^{n} f_i x_i}{f_i}$$

Łatwo jest wykazać, że licznik tego wzoru w zastosowaniu do funkcji de Moivre'a równa się zeru. Istotnie niech będzie *ABCD* jakikolwiek wycinek po prawej stronie krzywej funkcji (ryc. 6). Pole jego będzie równe częstości dla klasy *CD*, względnie dla jej wartości środkowej, określonej przez środek odcinka *CD*. Dla każdego wycinka z figury krzywej można znaleźć symetrycznie po drugiej stronie krzywej podobny wycinek *A'B'C'D'*, którego pole będzie reprezentowało częstość dla klasy *C'D'* z taką samą wartością środkową, ale ujemną. Mnożąc pola tych wycinków przez $+x$ w jednym przypadku, przez $-x$ w drugim, otrzymamy jednakowe iloczyny, ale

o przeciwnych znakach. W ten sposób otrzymamy wartość zero dla średniej arytmetycznej funkcji de Moivre'a.



Ryc. 6

Daleko bardziej zawiłe jest zagadnienie momentów. Dla obliczenia ich trzeba pole figury zakreślonej przez krzywą i oś odciętych rozbić gęsto ustawionymi rzędnymi na liczne wąskie wycinki. Trzeba następnie pola tych wycinków mnożyć przez odpowiednie potęgi odległości ich środków od rzędnej środkowej, reprezentującej średnią arytmetyczną. Trzeba wreszcie dodać te iloczyny i podzielić otrzymaną sumę przez pole figury, która reprezentuje sumę częstości wszystkich klas wariantów. To dzielenie jest zresztą zbędne, bo wspomniane pole jest równe jedności. Im więcej będzie wycinków, im one będą skutkiem tego węższe, tym dokładniejszy wypadnie wynik. Dokładne wartości można otrzymać tylko za pomocą rachunku całkowego.

Za pomocą tego rachunku otrzymuje się dla częściowych momentów pierwszego stopnia wartości:

$$m_{11} = -\sigma\sqrt{\frac{2}{\pi}},\ m_{12} = +\sigma\sqrt{\frac{2}{\pi}}$$

Rachunek całkowy daje dla momentu drugiego stopnia bardzo proste równanie $m_2 = \sigma^2$. Moment trzeciego stopnia jest

wobec symetrii krzywej równy zeru. Wreszcie moment czwartego stopnia równa się $m_4 = 3\,\sigma^4$.

Z tego wynika, że charakterystyczna dla funkcji de Moivre'a stała σ jest średnim odchyleniem tej funkcji. Oznaczenie jej tą samą grecką literą co odchylenie średnie jest zatem bardzo celowe. Wynika następnie, że wskaźnik skośności jest równy zeru, a wskaźnik skupienia wariantów równy 3.

Po tym długim, ale koniecznym wstępie przystępujemy do rozwiązania zagadnienia postawionego na początku ustępu. Chodzi o to, że szereg dwumianowy ze zwiększeniem n zbliża się do funkcji de Moivre'a. Przeprowadzenie ścisłego dowodu jest niemożliwe bez rachunku nieskończonościowego. Nietrudno jest natomiast sprawdzić omawiane twierdzenie, porównując z tą funkcją poszczególne szeregi. Weźmiemy naprzykład szeregi $(\frac{1}{2} + \frac{1}{2})^4$ i $(\frac{1}{2} + \frac{1}{2})^{16}$.

Porównanie, do którego przystępujemy, wymaga użycia dla szeregów dwumianowych i dla funkcji de Moivre'a tej samej zmiennej niezależnej. Funkcję weźmiemy w formie $\varphi_0(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$ nie zawierającej stałej σ. Zmienna ξ jest, jak widzieliśmy, określona przez równanie $x = \sigma\,\xi$. Ponieważ nadto początek tej zmiennej zbiega się ze średnią arytmetyczną funkcji, która jest równa zeru, wobec tego równanie określające zmienną ξ trzeba napisać w formie:

$$x = m + \sigma\,\xi.$$

Ponieważ dla szeregów dwumianowych $m = np$, $\sigma = \sqrt{npq}$, równanie powyższe przybierze formę:

$$x = np + \xi\,\sqrt{npq}$$

skąd

$$\xi = \frac{x - np}{\sqrt{nqp}}$$

A więc dla szeregu $(\frac{1}{2} + \frac{1}{2})^4$ będziemy mieli:

$$\xi = \frac{x - 4 \times {}^1/_2}{\sqrt{4 \times {}^1/_2 \times {}^1/_2}} = x - 2$$

dla szeregu zaś $(\frac{1}{2} + \frac{1}{2})^{16}$:

$$\xi = \frac{x - 16 \times {}^1/_2}{\sqrt{16 \times {}^1/_2 \times {}^1/_2}} = \frac{x - 8}{2}$$

Co się tyczy częstości w dwumianowych szeregach, nie trzeba będzie wprowadzać żadnych zmian, gdyż są to częstości względne, a więc suma ich równa się jedności, tak samo jak suma częstości w funkcji de Moivre'a formy φ_0 W ujęciu graficznym znaczy to, że suma pól prostokątów w histogramie szeregu równa się polu figury zakreślonej przez krzywą funkcji i przez oś odciętych.

Dla pierwszego z naszych szeregów otrzymamy w ten sposób zestawienie w tab. 25,2, dla drugiego w tab. 25,3.

Tabela 25,2

Szereg $(\frac{1}{2} + \frac{1}{2})^6$

Zmienna ewen-	x	=	0	1	2	3	4
tualna	ξ	=	—2	—1	0	+1	+2
Częstości w 16-tych							
częściach jedności			1	4	6	4	1

Dla porównania z funkcją $\varphi_0(\xi)$ trzeba będzie jeszcze przedstawić częstości w formie ułamków dziesiętnych a mianowicie z czterema znakami, tak bowiem są obliczone wartości funkcji w tabelach, z których będziemy korzystali.

Trzeba wreszcie obliczyć pola wycinków krzywej funkcji $\varphi_0(\xi)$ odpowiadające częstościom naszych szeregów dwumianowych, względnie prostokątom odnośnych histogramów.

Tabela 25,3

Szereg $(\frac{1}{2} + \frac{1}{2})^{16}$

Zmienna ewentualna x	ξ	Częstości wyrażone w $\frac{1}{2^{16}} = \frac{1}{65536}$ częściach jedności	Zmienna ewentualna x	ξ	Częstości wyrażone w $\frac{1}{2^{16}} = \frac{1}{65536}$ częściach jedności
0	−4.0	1	9	+0.5	11440
1	−3.5	16	10	+1.0	8008
2	−3.0	120	11	+1.5	4368
3	−2.5	560	12	+2.0	1820
4	−2.0	1820	13	+2.5	560
5	−1.5	4368	14	+3.0	120
6	−1.0	8008	15	+3.5	16
7	−0.5	11440	16	+4.0	1
8	0.0	12870			

Obliczenie pól wycinków z figury krzywej uskutecznimy za pomocą całki Laplace'a. Dla przeprowadzenia rachunków trzeba będzie wyznaczyć wartości zmiennej ξ dla rzędnych ograniczających wycinki, inaczej mówiąc wartości tej zmiennej dla boków prostokątów histogramu. To zadanie sprowadza się do wyznaczania granic klasowych przedziałów, co było już rozpatrywane w ust. 5.

Dla szeregu $(\frac{1}{2} + \frac{1}{2})^4$ te granice będą:

$$-2.5, \quad -1.5, \quad -0.5, \quad +0.5, \quad +1.5, \quad +2.5.$$

Dla nich z tabeli *B* w końcu książki znajdziemy wartości całki Laplace'a: 0.0063, 0.0668, 0.3085, 0.6915, 0.9332, 0.9937.

Odejmując te wartości kolejno od siebie, otrzymamy potrzebne nam pola wycinków figury krzywej $\varphi_0(\xi)$. Zestawiając je z częstościami szeregu, otrzymamy tabelę 25,4. Widzimy, że zgodność jest niezła mimo małej wartości n.

Tabela 25,4

Porównanie szeregu $(\frac{1}{2} + \frac{1}{2})^4$ z całką L a p l a c e'a.

ξ	Szereg	Całka	Różnica
−2.0	0.0625	0.0605	+0.0020
−1.0	0.2550	0.2417	+0.0083
0.0	0.3750	0.3830	−0.0080
+1.0	0.2500	0.2417	+0.0083
+2.0	0.0625	0.0605	+0.0020

Zgodność będzie lepsza dla szeregu $(\frac{1}{2} + \frac{1}{2})^{10}$. Ponieważ ten szereg, tak samo zresztą jak i poprzedni, jest symetryczny, wystarczy rozpatrzyć połowę jego. Granice prostokątów będą kolejno — 4.25, — 3.75, — 3.25, — 2.75,... Otrzymamy zestawienie następujące. (tab. 25,5).

Tabela 25,5

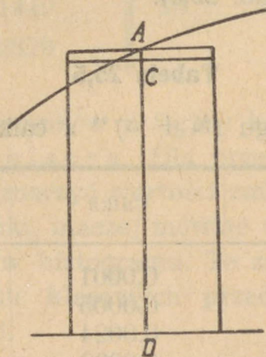
Porównanie szeregu $(\frac{1}{2} + \frac{1}{2})^{16}$ z całką L a p l a c e'a.

ξ	Szereg	Całka	Różnica
−4.0	0.0000	0.0001	−0.0001
−3.5	0.0002	0.0005	−0.0003
−3.0	0.0018	0.0024	−0.0006
−2.5	0.0085	0.0092	−0.0007
−2.0	0.0278	0.0279	−0.0001
−1.5	0.0666	0.0656	+0.0010
−1.0	0.1222	0.1200	+0.0012
−0.5	0.1746	0.1747	−0.0000
0.0	0.1964	0.1974	−0.0010

Dla jeszcze wyższych potęg n zgodność będzie jeszcze lepsza.

Porównanie szeregów dwumianowych z funkcją d e M o i v r e'a przeprowadziliśmy powyżej za pomocą całki L a-

p l a c e'a. Takie porównanie można także przeprowadzić bezpośrednio. Pola prostokątów histogramu pozostaną w dalszym ciągu odpowiednikami pól z figury krzywej $\varphi_0(\xi)$, ale dla tych ostatnich pól będziemy brali wartości przybliżone, a to w sposób następujący. Przez środek każdego wycinka przeprowadzimy rzędną tzw. r z ę d n ą ś r o d k o w ą (*AD* na ryc. 7) i pole wycinka przyrównamy polu prostokąta o tej samej podstawie z wysokością równą rzędnej środkowej. Popełni się przytem pewien błąd, który będzie tym mniejszy, im wycinek będzie węższy. Rzędna środkowa będzie się cokolwiek różniła od wysokości odnośnego prostokąta hisogramu (*CD* na ryc. 7), znowu tym mniej, im wycinek i prostokąt będą węższe. Teraz mamy do porównania zamiast wycinka i prostokąta — dwa prostokąty. Ponieważ te prostokąty mają wspólną podstawę, można porównywać nie ich pola, lecz wysokości, to znaczy rzędną środkową z wysokością prostokąta histogramu.



Ryc. 7

Dla przeprowadzenia tego porównania trzeba w wziąć pod uwagę, że przez wprowadzenie zmiennej ξ zamiast pierwotnej x została wprowadzona nowa jednostka pomiarowa dla podstaw prostokątów histogramu. Ta jednostka jest tyle razy większa od jedności, ile razy σ jest większe od jedności. Wynika to z równania zamiany zmiennej $\xi = \frac{x - m}{\sigma}$. Wobec tego podstawy prostokątów wyrażają się teraz liczbami mniejszymi niż było

poprzednio w stosunku $1 : \sigma$. Ażeby wielkość pól nie uległa zmianie, trzeba powiększyć w tym samym stosunku wysokości prostokątów, czyli pomnożyć częstości przez σ. Jeżeli więc oznaczymy częstości w szeregach przez y, będziemy mieli równanie zamiany zmiennej w formie $\eta = \sigma y$. Innymi słowami musimy tu wprowadzić zmienne normalne, takie, jakich użyliśmy na początku ustępu dla funkcji de Moivre'a.

Ponieważ dla pierwszego z rozpatrywanych szeregów dwumianowych $\sigma = 1$, częstości dla niego nie ulegną zmianie. Dla drugiego natomiast $\sigma = 2$, wypadnie wobec tego podwoić częstości. Biorąc z tabeli 25,1 wartości funkcji $\varphi_0(\xi)$, otrzymamy teraz porównanie zestawione w tabelach 25,6 i 25,7. Porównanie znowu wypadło lepsze dla większego n.

Tabela 25,6

Porównanie szeregu $(\frac{1}{2} + \frac{1}{2})^{16}$ z funkcją de Moivre'a.

ξ	Szereg	Funkcja	Różnice
−2.0	0.0625	0.0540	+0.0085
−1.0	0.2500	0.2420	+0.0080
0 .0	0.3750	0.3989	−0.0239

Tabela 25,7

Porównanie szeregu $(^1/_2 + {}^{10}/_2)^{16}$ z funkcją de Moivre'a.

ξ	Szereg	Funkcja	Różnice
−4.0	0.0000	0.0001	−0.0001
−3.5	0.0004	0.0009	−0.0005
−3.0	0.0036	0.0044	−0.0008
−2.5	0.0170	0.0175	−0.0005
−2.0	0.0556	0.0540	+0.0016
−1.5	0.1332	0.1295	+0.0037
−1.0	0.2444	0.2420	+0.0024
−0.5	0.3492	0.3521	−0.0029
0.0	0.3938	0.3989	−0.0001

Funkcja de Moivre'a i wyprowadzona z niej całka Laplace'a mają wielkie znaczenie w statystyce. Pochodzi to stąd, że liczne szeregi rozdzielcze mają rozdział częstości zbliżony do szeregu liczb, jakie otrzymuje się dla pól wycinków krzywej de Moivre'a za pomocą całki Laplace'a. Mówi się o takich szeregach, że mają normalny rozkład częstości albo że mają formę normalną. Już na oko widoczne to jest z przykładów przytoczonych w ust. 6. Dokładnie ta kwestia będzie omówiona w rozdziale VI. Jeszcze ważniejszym jest zastosowanie tej funkcji przy ocenie wiarogodności charakterystyk, co będzie tematem rozdziału V.

26. Szereg Poissona i jego charakterystyki. Widzieliśmy w ust. 23, że szereg dwumianowy wypada skośny, jeżeli p — prawdopodobieństwo zdarzenia — jest małe. Jeżeli jest bardzo małe, szereg staje się jednobocznym. Zwiększenie n, liczby przypadków w zbiorach, pomniejsza skośność.

Opierając się na powyższych zasadach, rozpatrzmy teraz zagadnienie powtarzania się zdarzenia, kiedy jego prawdopodobieństwo jest bardzo małe, a zbiory przypadków są bardzo duże. Przy takim założeniu iloczyn $\lambda = np$ oczywiście może mieć różne wartości, na ogół jednak ani zbyt duże, ani zbyt małe. Jaka będzie wtedy forma szeregu dwumianowego?

Otóż można dowieść, że będzie on miał formę zbliżoną do wzoru ustalonego przez francuskiego matematyka Poissona:

$$e^{-\lambda}\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots\right)$$

Z tego szeregu pierwszy wyraz, to znaczy $e^{-\lambda}$ daje prawdopodobieństwo niewystąpienia zdarzenia, drugi wyraz czyli $\lambda e^{-\lambda}$ — prawdopodobieństwo jednokrotnego wystąpienia zdarzenia, trzeci a więc $\frac{\lambda^2}{2!}e^{-\lambda}$ — prawdopodobieństwo dwukrotnego wystąpienia itd. Suma tych wyrazów jest równa jedności, gdyż suma wyrazów szeregu $1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots$

jest $= e^{+\lambda}$. Pod tym względem szereg Po'issona jest podobny do szeregu dwumianowego oraz do całki Laplace'a. Nietrudno jest przekonać się, że jeżeli λ jest mniejsze od jedności, to wyrazy szeregu będą malały. Naprzykład przy $\lambda = 0.5$ wypadnie (w tysiącznych):

$$607,\ 303,\ 76,\ 13,\ 2, \ldots$$

Przy $\lambda = 1$ pierwsze dwa wyrazy będą największe i będą sobie równe a dalsze coraz mniejsze: 368, 368, 184, 61, 15, 3;... Przy $\lambda > 1$ będą one z początku wzrastały a potem malały. Naprzykład przy $\lambda = 2$ będziemy mieli (zawsze w tysięcznych częściach jedności): 135, 271, 271, 180, 90, 36;...

Widzimy w tym ostatnim przykładzie, że w nim tak samo jak w poprzednim dwa wyrazy są jednakowe, tylko nie pierwszy i drugi, lecz drugi i trzeci. Przy całkowitych wartościach λ zawsze dwa kolejne wyrazy wypadną jednakowe. Będą to wyrazy największe. Tego nie będzie przy niecałkowitych wartościach λ, naprzykład przy $\lambda = 2.5$ otrzymamy szereg: 82, 205, 257, 214, 134, 67,...

Szereg Poissona jest nieskończony, ale z niego trzeba wziąć tylko $n + 1$ wyrazów, tyle ma ich bowiem szereg dwumianowy. Stąd wynika pewna niezgodność między tymi szeregami. Ponieważ jednak wyrazy szeregu Poissona maleją bardzo szybko, a liczba n jest bardzo duża, różnica między omawianymi szeregami jest minimalna, co można łatwo sprawdzić.

I tak przypuśćmy, że prawdopodobieństwo pewnego zdarzenia jest jedna setna a zbiory liczą po sto przypadków. Będziemy wtedy mieli $\lambda = np = 1$ i szereg Poissona przybierze formę $\frac{1}{e}(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \ldots) + \frac{1}{120} + \ldots)$

Odpowiedni szereg dwumianowy będzie następujący:

$$\left(\frac{99}{100} + \frac{1}{100}\right)^{100} = \left(\frac{99}{100}\right)^{100} + \frac{100}{1}\left(\frac{99}{100}\right)^{99}\frac{1}{100} +$$

$$+\frac{100.99}{1.2}\left(\frac{99}{100}\right)^{98}\frac{1}{100^2}+\frac{100.99.98}{1.2.3}\left(\frac{99}{100}\right)^{97}\frac{1}{100^3}+\ldots$$

Porównanie kolejnych wyrazów tych dwóch szeregów da wynik następujący w tysięcznych częściach jedności (tab. 26,1).

Tabela 26,1

Szereg Poissona	Szereg dwumianowy	Różnica
368	367	+1
368	370	−2
184	185	−1
61	61	0
15	15	0
3	3	0

Zgodność wypadła doskonała.

Trzeba jeszcze obliczyć średnią arytmetyczną i momenty dla szeregu Poissona. Dla obliczenia tych charakterystyk weźmiemy za wartość wyjściową zero. Dla średniej arytmetycznej będziemy mieli (tab. 26, 2). Sumowanie iloczynów fx da nam po wyprowadzeniu za nawias czynnika $\lambda e^{-\lambda}$:

$$\Sigma fx = \lambda e^{-\lambda}\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\ldots\right)$$

Ponieważ suma szeregu stojącego w nawiasach wynosi e^{λ}, suma odchyleń będzie równa λ. Taką samą wartość będzie miał pomocniczy moment μ_1, gdyż suma f równa się jedności. Ponieważ średnia arytmetyczna równa się μ_1 przy wartości wyjściowej zero, będzie ona także równa λ.

Tabela 26,2

x Liczby powtórzeń zdarzenia	f Częstość względna tych liczb	fx
0	$e^{-\lambda}$	0
1	$\lambda e^{-\lambda}$	$\lambda e^{-\lambda}$
2	$\frac{\lambda^2}{1.2}e^{-\lambda}$	$\lambda^2 e^{-\lambda}$
3	$\frac{\lambda^3}{1.2.3}e^{-\lambda}$	$\frac{\lambda^3}{1.2}e^{-\lambda}$
4	$\frac{\lambda^4}{1.2.3.4}e^{-\lambda}$	$\frac{\lambda^4}{1.2.3}e^{-\lambda}$

Momentów pierwszego stopnia nie można określić w ogólnej formie dla tych samych powodów, jak dla szeregu dwumianowego. Natomiast moment drugiego stopnia da się obliczyć z łatwością tą samą metodą. Będziemy mieli według zestawienia tabeli 26,3:

$$\mu_2 = \lambda e^{-\lambda}\left(1 + 2\lambda + 3\frac{\lambda^2}{1.2} + 4\frac{\lambda^3}{1.2.3} + 5\frac{\lambda^4}{1.2.3.4} + \ldots\right)$$

Wydzielając z wielomianu stojącego w nawiasach wielomian

$$1 + \lambda + \frac{\lambda^2}{1.2} + \frac{\lambda^3}{1.2.3} + \ldots = e^{\lambda}$$

otrzymamy: $\mu_2 = \lambda e^{-\lambda}\left(e^{\lambda} + \lambda + 2\frac{\lambda^2}{1.2} + 3\frac{\lambda^3}{1.2.3} + 4\frac{\lambda^4}{1.2.3.4} + \right.$

$$\left. + \ldots\right) = \lambda e^{-\lambda}\left[e^{\lambda} + \lambda\left(1 + \frac{\lambda}{1} + \frac{\lambda^2}{1.2} + \frac{\lambda^3}{1.2.3} + \ldots\right)\right] =$$

$$= \lambda e^{-\lambda}\left(e^{\lambda} + \lambda e^{\lambda}\right) = \lambda + \lambda^2$$

Stąd wypadnie $m_2 = \mu_2 - \mu_1^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$

W podobny sposób można wyprowadzić wzory dla momentów trzeciego i czwartego stopnia: $m_3 = \lambda$, $m_4 = \lambda + 3\lambda^2$.

Tabela 26,3

x	f	x^2	xf^2
0	$e^{-\lambda}$	0	0
1	$\lambda e^{-\lambda}$	1	$\lambda e^{-\lambda}$
2	$\frac{\lambda^2}{1\cdot 2} e^{-\lambda}$	4	$2 \frac{\lambda^2}{1} e^{-\lambda}$
3	$\frac{\lambda^3}{1\cdot 2\cdot 3} e^{-\lambda}$	9	$3 \frac{\lambda^3}{1\cdot 2} e^{-\lambda}$
4	$\frac{\lambda^4}{1\cdot 2\cdot 3\cdot 4} e^{-\lambda}$	16	$4 \frac{\lambda^4}{1\cdot 2\cdot 3} e^{-\lambda}$
5	$\frac{\lambda^5}{1\cdot 2\cdot 3\cdot 4\cdot 5} e^{-\lambda}$	25	$5 \frac{\lambda^5}{1\cdot 2\cdot 3\cdot 4} e^{-\lambda}$

Szereg Poissona jest bardzo pożyteczny przy opracowywaniu jednobocznych szeregów rozdzielczych.

ROZDZIAŁ V.

WIARYGODNOŚĆ CHARAKTERYSTYK MATERIAŁÓW STATYSTYCZNYCH.

27. Istota zagadnienia wiarogodności charakterystyk. Omówione w rozdziałach II i III charakterystyki materiałów statystycznych były ustalone bez względu na to czy te materiały obejmowały całą populację, czy też stanowiły jej część. Otóż zdarza się tylko wyjątkowo, żeby opracowywano w całości materiał jakiegoś zagadnienia, nawet jeżeli jest on ograniczony i dostępny, jak na przykład spis ludności. Prawie zawsze bierze się z populacji tylko niewielką jej część, wybraną umyślnie w sposób przypadkowy i na podstawie tej tak zwanej próby sądzi się o całości populacji. Naturalnie tylko wyjątkowo może się zdarzyć, że próba jest w zupełności podobna do populacji. Stajemy więc wobec niezmiernie ważnego zagadnienia — w jakim stopniu próba jest podobna do populacji. W tłumaczeniu na język statystyki będzie to równoważne z pytaniem, w jakim stopniu charakterystyki próby zbliżają się do takich samych charakterystyk całości populacji. Jest to zagadnienie wiarogodności tych charakterystyk.

Bezpośredniej odpowiedzi na powyższe pytanie dać nie można. Byłoby to bowiem równoznaczne z dokładną charakterystyką populacji według próby, co jest niemożliwe. Można tylko ustalić, jaka część wszelkich możliwych prób ma taką czy inną zgodność z populacją.

Inaczej można powiedzieć posługując się językiem nauki o prawdopodobieństwie, że *określenie wiarogodności charakterystyk może polegać tylko na ustaleniu prawdopodobieństwa takiego lub innego stopnia zgodności między próbą a populacją.*

Trzeba zaraz na wstępie zaznaczyć, że pojęcie prawdopodobieństwa będzie tu odmienne od podanego w poprzednim rozdziale. Różnica nie jest zasadnicza, — inaczej nie można by było używać tego samego terminu. I tu i tam chodzi o częstość zdarzeń, o ile ona jest powodowana przez określone przyczyny. Ale przyczyny określone sprowadzają się teraz do jednej tylko — tej samej we wszystkich zagadnieniach — do składu populacji. I to jeszcze nie jest rzecz główna — ta przyczyna jest zawsze okryta tajemnicą, podczas gdy w nauce o prawdopodobieństwie przyczyny określone były znane. Tam mogliśmy za pomocą kombinatoryki z tych przyczyn wyprowadzić liczbową wartość prawdopodobieństwa. Tu jest to niemożliwe. Mamy tu prawdopodobieństwo osobliwego rodzaju — *prawdopodobieństwo statystyczne.*

A więc stajemy przed zadaniem nierozwiązalnym — powie czytelnik. I tak, i nie. Zadanie da się rozwiązać częściowo a to w sposób następujący. Zakładamy, że populacja ma taki a taki skład i badamy, jak się będą przedstawiały wszelkie możliwe próby, jakie z niej można otrzymać. Ustalamy tą drogą związki zachodzące między populacją a ogółem prób. Poznawszy te związki, będziemy mogli w pewnej mierze sądzić o nieznanym charakterze populacji według wziętych z niej prób. Zazwyczaj zakłada się, że populacja ma charakter normalny. Dodatkowo bada się, o ile odchylenia od tego charakteru wpływają na związek między populacją a próbami.

Z powyższego wypływa, że w zagadnieniach wiarogodności charakterystyk prowadzi się rozumowania tymi samymi metodami co w nauce o prawdopodobieństwie. Stąd też jej podstawowe twierdzenia, w szczególności twierdzenie o prawdopodobieństwie całkowitym i złożonym, zachowują swoją ważność. Można by nawet ogólne pojęcie prawdopodobieństwa ująć w spo-

sób statystyczny. Można mianowicie rozpatrywane zbiory zdarzeń traktować jako próby wyjęte z pewnego rodzaju populacji obejmującej wszystkie możliwe zbiory danych zdarzeń. Wtedy ilościowe określenie prawdopodobieństwa sprowadziłoby się po prostu do względnej częstości zdarzenia w takiej populacji. Tak też traktuje się prawdopodobieństwo w statystyce. O ile się w niej mówi o prawdopodobieństwie, chodzi po prostu o względną częstość w populacji. Trzeba przy tym operować uważnie takimi pojęciami, mianowicie trzeba uważać na to, by nie pomieszać populacji z próbą. Częstość względna wariantów w próbie nie może być uważana za żaden rodzaj prawdopodobieństwa.

W związku z powyższym trzeba jeszcze bliżej określić pojęcie próby. Traktowaliśmy ją jako część populacji. Takie określenie jest niedostateczne. Dotyczy ono jednej tylko odmiany omawianego pojęcia — próby bez powtórzeń. A jest jeszcze druga odmiana: próba z powtórzeniami.

O co tu chodzi? Jeżeli próba ma być częścią populacji, to wszystkie zawarte w niej warianty muszą być różne. Niektóre z nich mogą być jednakowe, ale nie te same. Na przykład jeżeli z urny zawierającej białe i czarne kule wybierzemy część ich, powiedzmy dziesięć, to może w niej być sześć jednakowych czarnych kul i cztery jednakowe białe, ale będą to kule różne. Można jednak utworzyć próbę w inny sposób. Można mianowicie wziąć z urny jedną kulę, zanotować jej barwę i włożyć ją z powrotem. Następnie po wymieszaniu można wziąć — zawsze na chybi trafi — znowu jedną z kul i po zanotowaniu jej barwy włożyć ją z powrotem do urny. Powtarzając tę czynność dziesięć razy, możemy otrzymać próbę złożoną także z sześciu czarnych i czterech białych kul, ale o innym charakterze. W niej mogą się spotkać kule nietylko podobne ale i identyczne — może się przecież zdarzyć, że wyciągniemy tę samą kulę nawet wielokrotnie. Będzie to próba z powtórzeniami. Natomiast wyjęcie dziesięciu kul z urny bez wkładania ich z powrotem da próbę bez powtórzeń.

W praktyce statystycznej ma się do czynienia zwykle z próbami bez powtórzeń. Istotnie, jeżeli na przykład liczy się kwiaty w koszykach bławatka, to bierze się coraz inny koszyk, albo jeżeli mierzy się wzrost poborowych, to ma się do czynienia z coraz to innym osobnikiem. Jednakże przy wielkich w porównaniu z próbami populacjach, a tym bardziej przy nieograniczonych, branie prób bez powtórzeń daje ten sam wynik co z powtórzeniami. Istotnie usuwa się przy tym pewne warianty z populacji, ale wobec wielkiej ich ilości skład populacji zachowuje prawie ten sam charakter przy ograniczonym jej zakresie i dokładnie ten sam przy nieograniczonym. Jest wtedy bardzo mało prawdopodobne, by usunięcie z populacji składników jednej próby miało wpłynąć na skład dalszych prób.

Zagadnienia, które będziemy omawiali w tym rozdziale, muszą być traktowane inaczej przy próbach z powtórzeniami, aniżeli przy próbach bez powtórzeń. Rozumowania i wyniki w pierwszym przypadku są o wiele prostsze, a ponieważ w praktyce różnica między jednym a drugim rodzajem prób zaciera się, będziemy uważali, że rozpatrywane próby są brane z powtórzeniami, jakkolwiek w rzeczywistości jest odwrotnie.

W dalszych ustępach będziemy badali przedstawione tu zagadnienie wiarogodności. Jest to zadanie zawiłe i trudne. Trzeba będzie rozwiązywać je stopniowo, przy czym wypadnie wprowadzać nowe pojęcia i stosować naukę o prawdopodobieństwie. Narazie można powiedzieć ogólnie, że im próba jest większa, tym bardziej zbliża się swoim charakterem do populacji, ale niekoniecznie. Przypadek może zrządzić, że właśnie wielka próba będzie się różniła silnie od populacji. Trafia się to tym rzadziej, im próba jest większa, ale trafić się może zawsze.

Zwiększenie zakresu prób nie rozwiązuje zatem kwestii. Nadto otrzymywanie wielkich prób jest trudne i często nawet praktycznie niewykonalne. Nieraz wypada pracować z próbami po kilka wariantów zaledwie, jak to jest prawidłem w doświadczalnictwie rolniczym. Zagadnienie małych prób ma przeto szczególne znaczenie w statystyce i będzie omówione osobno w ustępie 31.

28. Populacje charakterystyk. Z wywodów poprzedniego ustępu wynika, że zagadnienie wiarogodności charakterystyk materiałów statystycznych wymaga zbadania ogółu prób, jakie mogą być otrzymane z populacji. Każda próba ma swoje charakterystyki: średnią, momenty itd. Możemy wobec tego utworzyć dla każdej charakterystyki osobne populacje, tzw. *wtórne*, złożone z wartości charakterystyk poszczególnych prób. Dla uniknięcia nieporozumień można populację, z której się one wywodzą, nazywać pierwotną. Populacje wtórne są zawsze nieograniczone, nawet jeżeli populacja pierwotna jest ograniczona. Można bowiem bez końca wybierać próby z populacji, jak kule z urny, jeżeli się je wkłada z powrotem. Będą to naturalnie próby z powtórzeniami.

Populacje wtórne mają swoje charakterystyki, jak każda populacja. Nadano im groteskową nazwę „nadziei matematycznych". A więc przede wszystkim będziemy mieli średnią arytmetyczną, która odgrywa szczególnie ważną rolę.

Dalej ważną charakterystyką wtórnych populacji jest moment drugiego stopnia, względnie średnie odchylenie, będące pierwiastkiem kwadratowym z tego momentu. Stwarza on podstawę do obliczenia tak zwanego *średniego błędu*, którym są obciążone charakterystyki prób.

Wreszcie nie można pominąć momentów trzeciego i czwartego stopnia, względnie wyprowadzanych z nich pochodnych charakterystyk — skośności i ekscesu. Co do tych ostatnich zachodzi rzecz wielkiej wagi: *przy zwiększeniu zakresu prób skośność i eksces populacji wtórnych zmierzają do zera*. To znaczy, że *przy zwiększeniu zakresu prób populacje wtórne zbliżają się do formy normalnej*. Trudności matematyczne nie pozwalają podać dowodu tego podstawowego twierdzenia. Można jednak dla wyjaśnienia istoty zagadnienia sprawdzić je na konkretnych przykładach. Zróbmy to dla średniej arytmetycznej.

Pierwszym takim przykładem będzie bardzo prosta populacja złożona z 5 wariantów: 1, 2, 3, 4, 5. Weźmiemy z niej próby po trzy warianty naturalnie z powtórzeniami i rozpa-

trzymy populację wtórną złożoną ze średnich arytmetycznych dla tych prób. Pomimo nieograniczności tej popupalcji możemy ustalić jej skład. Istotnie łatwo jest ustalić listę równie prawdopodobnych spotkań wariantów przy wybieraniu prób. Będzie to lista obejmująca $5^3 = 125$ takich spotkań (tab. 28,1).

Tabela 28,1

1 1 1	1 1 2	1 1 3	1 1 4	1 1 5
1 2 1	1 2 2	1 2 3	1 2 4	1 2 5
1 3 1	1 3 2	1 3 3	1 3 4	1 3 5
1 4 1	1 4 2	1 4 3	1 4 4	1 4 5
1 5 1	1 5 2	1 5 3	1 5 4	1 5 5
2 1 1	2 1 2	2 1 3	2 1 4	2 1 5
2 2 1	2 2 2	2 2 3	2 2 4	2 2 5
2 3 1	2 3 2	2 3 3	2 3 4	2 3 5
2 4 1	2 4 2	2 4 3	2 4 4	2 4 5
2 5 1	2 5 2	2 5 3	2 5 4	2 5 5
3 1 1	3 1 2	3 1 3	3 1 4	3 1 5
3 2 1	3 2 2	3 2 3	3 2 4	3 2 5
3 3 1	3 3 2	3 3 3	3 3 4	3 3 5
3 4 1	3 4 2	3 4 3	3 4 4	3 4 5
3 5 1	3 5 2	3 5 3	3 5 4	3 5 5
4 1 1	4 1 2	4 1 3	4 1 4	4 1 5
4 2 1	4 2 2	4 2 3	4 2 4	4 2 5
4 3 1	4 3 2	4 3 3	4 3 4	4 3 5
4 4 1	4 4 2	4 4 3	4 4 4	4 4 5
4 5 1	4 5 2	4 5 3	4 5 4	4 5 5
5 1 1	5 1 2	5 1 3	5 1 4	5 1 5
5 2 1	5 2 2	5 2 3	5 2 4	5 2 5
5 3 1	5 3 2	5 3 3	5 3 4	5 3 5
5 4 1	5 4 2	5 4 3	5 4 4	5 4 5
5 5 1	5 5 2	5 5 3	5 5 4	5 5 5

W tej liście część prób różni się między sobą tylko porządkiem wariantów. Zbierając razem podobne próby, otrzymamy 35 rodzajów ich. Częstość tych rodzajów jest częściowo różna. Nietrudno jest ustalić te częstości. Będziemy mieli nową listę prób z częstościami względnymi w 125-tych częściach jedności (tab. 28,2). Te częstości będą oczywiście niczym innym, jak prawdopodobieństwami statystycznymi. We wspomnianej tabeli są podane oprócz tych prawdopodobieństw średnie arytmetyczne wyrażone w trzecich częściach jedności. Na podstawie tych danych można teraz ułożyć szereg rozdzielczy wtórnej populacji, złożonej ze średnich dla prób (tab. 28,3).

Szereg ten wypadł symetryczny — tak jest zawsze, jeżeli populacja wtórna wywodzi się z populacji symetrycznej. Co się zaś tyczy ekscesu, to wypada on ujemny i równa się —0.538. Widzimy, że rozpatrywana populacja nie odbiega zbytnio od normalnej.

Tabela 28,2

Rodzaje prób	Ich prawdopodobieństwa w $^1/_{125}$	Ich średne w ⅓
1 1 1	1	3
1 1 2	3	4
1 1 3	3	5
1 1 4	3	6
1 1 5	3	7
1 2 2	3	5
1 2 3	6	6
1 2 4	6	7
1 2 5	6	8
1 3 3	3	7
1 3 4	6	8
1 3 5	6	9

Rodzaje prób	Ich prawdopodobieństwa w $^1/_{125}$	Ich średnie w $\frac{1}{3}$
1 4 4	3	9
1 4 5	6	10
1 5 5	3	11
2 2 2	1	6
2 2 3	3	7
2 2 4	3	8
2 2 5	3	9
2 3 3	3	8
2 3 4	6	9
2 3 5	6	10
2 4 4	3	10
2 4 5	6	11
2 5 5	3	12
3 3 3	1	9
3 3 4	3	10
3 3 5	3	11
3 4 4	3	11
3 4 5	6	12
3 5 5	3	13
4 4 4	1	12
4 4 5	3	13
4 5 5	3	14
5 5 5	1	15

Ale tego jest mało. Trzeba wziąć próby większe.

Weźmiemy w tym celu inną populację, umyślnie silnie asymetryczną, nawet jednoboczną. Skład jej będzie następujący: 50 % wariantów z wartością 1, 30 % wariantów z wartością 2 i 20 % wariantów 3. Inaczej można powiedzieć, że statystyczne prawdopodobieństwo wariantów 1 jest 0.5, wariantów 2 jest 0.3 i wariantów 3 jest 0.2.

Tabela 28,3

Średnie prób w $^1/_3$	Ich prawdopodobieństwa w $^1/_{125}$
3	1
4	3
5	6
6	10
7	15
8	18
9	19
10	18
11	15
12	10
13	6
14	3
15	1

Weźmiemy z początku próby po 3 warianty, tak jak poprzednio. Będziemy mieli tylko 10 rodzajów prób: 111, 112, 113, 122, 123, 133, 222, 223, 233, 333. Częstości względne tych rodzajów prób, inaczej mówiąc ich prawdopodobieństwa statystyczne, są różne. Dadzą się one obliczyć z łatwością według uogólnionego dwumianu Newtona (ust. 23). Będziemy mieli prawdopodobieństwa (tab. 28,4).

Tabela 28,4

Prawdopodobieństwa

dla prób 1 1 1 $\frac{3!}{3!\,0!\,0!} \times 0.5^3 \times 0.3^0 \times 0.2^0 = 0.125$

„ „ 1 1 2 $\frac{3!}{2!\,1!\,0!} \times 0.5^2 \times 0.3^1 \times 0.2^0 = 0.225$

„ „ 1 1 3 $\frac{3!}{2!\,0!\,1!} \times 0.5^2 \times 0.3^0 \times 0.2^1 = 0.150$

„ „ 1 2 2 $\frac{3!}{1!\,2!\,0!} \times 0.5^1 \times 0.3^2 \times 0.2^0 = 0.135$

„ „ 1 2 3 $\frac{3!}{1!\,1!\,1!} \times 0.5^1 \times 0.3^1 \times 0.2^1 = 0.180$

„ „ 1 3 3 $\frac{3!}{1!\,0!\,2!} \times 0.5^1 \times 0.3^0 \times 0.2^2 = 0.060$

„ „ 2 2 2 $\frac{3!}{0!\,3!\,0!} \times 0.5^0 \times 0.3^3 \times 0.2^0 = 0.027$

„ „ 2 2 3 $\frac{3!}{0!\,2!\,1!} \times 0.5^0 \times 0.3^2 \times 0.2^1 = 0.054$

„ „ 2 3 3 $\frac{3!}{0!\,1!\,2!} \times 0.5^0 \times 0.3^1 \times 0.2^2 = 0.036$

„ „ 3 3 3 $\frac{3!}{0!\,0!\,3!} \times 0.5^0 \times 0.3^0 \times 0.2^3 = 0.008$

Na tej podstawie możemy ułożyć szereg rozdzielczy średnich dla prób trójkowych (tab. 28,5). Wypadł szereg wyraźnie skośny.

Tabela 28,5

Próby trójkowe

Średnie w ⅓	Ich prawdopodobieństwa
3	0.125
4	0.225
5	0.285
6	0.207
7	0.114
8	0.036
9	0.008

Układając tą samą metodą szeregi dla prób większych — czwórkowych, piątkowych i szóstkowych — otrzymamy szeregi mniej skośne, pomimo skrajnej skośności rozpatrywanej pierwotnej populacji (tab. 28,6—8).

Tabela 28,6

Próby czwórkowe

Średnie w ¼	Ich prawdopodobieństwa
4	0.0625
5	0.1500
6	0.2350
7	0.2340
8	0.1761
9	0.0936
10	0.0376
11	0.0096
12	0.0016

Tabela 28,7

Próby piątkowe

Średnie w $^1/_5$	Ich prawdopodobieństwa
5	0.03125
6	0.09375
7	0.17500
8	0.21750
9	0.20525
10	0.14643
11	0.08210

Średnie w $^1/_5$	Ich prawdopodobieństwa
12	0.03480
13	0.01120
14	0.00240
15	0.00032

Tabela 28,8

Próby szóstkowe

Średnie w $^1/_6$	Ich prawdopodobieństwa
6	0.015625
7	0.056250
8	0.121875
9	0.180000
10	0.202875
11	0.178290
12	0.126029
13	0.071316
14	0.032460
15	0.011520
16	0.003120
17	0.000576
18	0.000064

Porównajmy ostatni, szóstkowy, z tych szeregów z krzywą normalną. Weźmiemy do tego cztery znaki dziesiętne zamiast wszystkich sześciu w wartościach prawdopodobieństw. Obliczmy odchylenie średnie. Otrzymamy $\varrho = 1.9136$, wyrażone w $^1/_6$ jako jednostce. Stąd wyprowadzimy za pomocą całki Laplace'a następujące porównanie populacji z krzywą normalną (tab. 28,9).

Tabela 28,9

Próby szóstkowe

Średnie w $^1/_6$	Ich prawdopodobieństwa w $^1/10000$	Wartości krzywej normalnej w $^1/10000$	Różnice
6	156	198	−42
7	562	525	+37
8	1219	1074	+145
9	1800	1690	+110
10	2029	2079	−50
11	1783	1881	−98
12	1260	1371	−111
13	713	685	+28
14	325	305	+20
15	115	94	+21
16	31	23	+8
17	6	4	+2
18	1	1	0

Widzimy, że zgodność wypadła niezła, jakkolwiek próby nie były duże.

29. Nadzieja matematyczna. Jak to było podane w poprzednim ustępie, nadzieja matematyczna takiej czy innej charakterystyki jest to średnia arytmetyczna wartości danej charakterystyki dla ogółu prób wziętych z danej populacji, inaczej mówiąc, średnia arytmetyczna odnośnej wtórnej populacji. Oznacza się ją literą *E* (espérance) z podaniem w nawiasach symbolu odnośnej charakterystyki. A więc na przykład *E (m)* jest nadzieją matematyczną średniej arytmetycznej.

Znaczenie nadziei matematycznej wynika stąd, że daje ona pewne pojęcie o wartości charakterystyk dla populacji, daje nadzieję bliższego oznaczenia takich charakterystyk na pod-

stawie prób z populacji wziętych. Jak każda nadzieja, jest ona złudna i zbyt wiele od niej spodziewać się nie można. Sprawa przedstawia się w sposób następujący. Charakterystyki nieograniczonej populacji nie mogą być obliczone bezpośrednio. Chcemy jednak je ustalić przynajmniej w przybliżeniu. Otóż dla każdej charakterystyki populacji można ustalić zupełnie dokładnie, za pomocą równania, związek jej z odnośną nadzieją matematyczną. Zadanie sprowadza się zatem do obliczenia tej nadzieji. Ponieważ jest to średnia arytmetyczna wartości charakterystyki dla prób, można ją obliczyć z dowolną dokładnością, jeżeli się weźmie tych prób dostatecznie dużo. W praktyce sprawa przedstawia się jednak niekorzystnie, gdyż ma się zwykle jedną tylko próbę. Widoki należytego rozwiązania poprawiają się, jeżeli próba jest duża, ale zasadnicza niepewność pozostaje, gdyż można przypadkowo natrafić na próbę znacznie różniącą się od populacji. Im próba jest większa, tym rzadziej taki „pech" wystąpi.

Rozpatrzmy nadzieje matematyczne poszczególnych charakterystyk. Wartości charakterystyk dla prób będziemy oznaczali małymi literami, tak jak dotąd, natomiast ich wartości dla populacji — przez odnośne duże litery, na przykład średnią arytmetyczną próby będziemy oznaczali przez m, średnią zaś populacji przez M. Będziemy mieli tu do czynienia z równaniami, których wyprowadzenie jest bardzo zawiłe. To też muszę ograniczyć się do podania tych równań i sprawdzenia ich ważności na przykładach.

Najprościej przedstawia się rozpatrywane zagadnienie dla średniej arytmetycznej: n a d z i e j a m a t e m a t y c z n a ś r e d n i e j r ó w n a s i ę ś r e d n i e j d l a p o p u l a c j i: $E(m) = M$. Nie trudno jest sprawdzić to twierdzenie na przykładach poprzedniego ustępu. Dla populacji 1, 2, 3, 4, 5 $M = 3$. Z podanego w tabeli 28,3 szeregu rozdzielczego średnich dla ogółu prób wynika nawet bez obliczeń taka sama wartość.

Dla drugiej z rozpatrywanych w poprzednim ustępie populacyj $M = 1.7$. Obliczmy średnią dla wtórnej populacji zło-

żonej z trójkowych prób. Na podstawie danych tabeli 28,5 otrzymamy:

$$E(m) = \frac{\Sigma f x}{\Sigma f} = {}^1/_3 (3 \times 0.125 + 4 \times 0.225 + 5 \times 0.285 +$$

$$+ 6 \times 0.207 + 7 \times 0.114 + 8 \times 0.036 + 9 \times 0.008) : (0.125 +$$

$$+ 0.225 + 0.285 + 0.207 + 0.114 + 0.036 + 0.008) =$$

$$= {}^1/_3 (0.375 + 0.900 + 1.425 + 1.242 + 0.798 + 0.288 +$$

$$+ 0.072) = {}^1/_3 \times 5.100 = 1.7 = M$$

Czytelnik może łatwo przekonać się, że taki sam wynik dadzą wtórne populacje większych prób tej populacji.

Przechodzimy do momentu drugiego stopnia. Tu nadzieja matematyczna nie będzie równa wartości tego momentu dla populacji. Będzie mianowicie trochę za mała:

$$E(m_2) = \frac{n-1}{n} M_2$$

W tym równaniu n, jak zawsze, jest liczbą wariantów w próbie.

Łatwo jest sprawdzić to równanie na konkretnym przykładzie, np. na populacji 1, 2, 3, 4, 5. Dla niej $M_2 = 2$. Obliczmy nadzieję matematyczną według danych tabeli 28,2. Tabele 29,1 i 29,2 dają przebieg obliczeń.

Tabela 29,1

Rodzaje prób	Ich prawdopodobieństwa w $^1/125$	Ich momenty drugiego stopnia w $^1/27$	Rodzaje prób	Ich prawdopodobieństwa w $^1/125$	Ich momenty drugiego stopnia w $^1/27$
1 1 1	1	0	1 3 3	3	24
1 1 2	3	6	1 3 4	6	42
1 1 3	3	24	1 3 5	6	72
1 1 4	3	54	1 4 4	3	54
1 1 5	3	96	1 4 5	6	78
1 2 2	3	6	1 5 5	3	96
1 2 3	6	18	2 2 2	1	0
1 2 4	6	42	2 2 3	3	6
1 2 5	6	78	2 2 4	3	24

Rodzaje prób	Ich prawdopodobieństwa w $^1/125$	Ich momenty drugiego stopnia w $^1/125$	Rodzaje prób	Ich prawdopodobieństwa w $^1/125$	Ich momenty drugiego stopnia w $^1/27$
2 2 5	3	54	3 3 5	3	24
2 3 3	3	6	3 4 4	3	6
2 3 4	6	18	3 4 5	6	18
2 3 5	6	42	3 5 5	3	24
2 4 4	3	24	4 4 4	1	0
2 4 5	6	42	4 4 5	3	6
2 5 5	3	54	4 5 5	3	6
3 3 3	1	0	5 5 5	1	0
3 3 4	3	6			

Tabela 29,2

Wartości momentów drugiego stopnia w $^1/27$	Ich prawdopodobieństwa w $^1/125$	Wartości momentów drugiego stopnia w $^1/27$	Ich prawdopodobieństwa w $^1/125$
0	5	54	12
6	24	72	6
18	18	78	12
24	18	96	6
42	24		

Wypadnie ostatecznie średnia arytmetyczna wtórnej populacji złożonej z momentów drugiego stopnia dla prób:

$$E(m_2) = \frac{4}{3},$$

podczas gdy moment drugiego stopnia dla pierwotnej populacji wynosi $M_2 = 2$. Ponieważ n w tym wypadku równa się 3, więc zgodnie z rozpatrywanym równaniem otrzymamy:

$$E(m_2) = \frac{3-1}{3} \times 2 = \frac{4}{3}$$

Z powyższego wynika prawidło do obliczenia momentów drugiego stopnia dla prób. Zamiast tego, żeby dzielić sumę kwa-

dratów odchyleń przez liczbę wariantów, trzeba ją dzielić przez liczbę o jedynkę mniejszą według wzoru

$$m_2 = \frac{\Sigma (u_i - \bar{u})^2}{n-1}$$

W ten sposób otrzymuje się wartości bliższe wartości tego momentu dla populacji. Naturalnie ma to znaczenie tylko przy małych próbach, ale z takimi próbami właśnie wypada często pracować.

Obliczenie średniego odchylenia dla prób ulegnie odpowiedniej zmianie. Dla uniknięcia nieporozumień dobrze jest otrzymane tą zmienioną metodą wartości oznaczać nie grecką literą σ lecz odnośną łacińską s

$$s = \sqrt{\frac{\Sigma (u_i - \bar{u})}{n-1}} = \sigma \sqrt{\frac{n}{n-1}}$$

Nadzieje matematyczne momentów wyższych stopni również nie są równe wartościom tych momentów dla populacji. Mianowicie mamy dla nich równania

$$E(m_3) = \frac{(n-1)(n-2)}{n^2} M_3$$

$$E(m_4) = \frac{(n-1)(n-2)(n-3)}{n^3} M_4 + \frac{(n-1)(2n-3)}{n^3} M_4 + 3M^2_2$$

Te równania nie mają praktycznego znaczenia.

30. Błąd średni i prawdopodobny. Wywody poprzednich ustępów miały charakter wstępu do właściwego zagadnienia wiarogodności charakterystyk. Rozwiązanie tego zagadnienia może być częściowo przeprowadzone za pomocą pojęcia *błędu średniego*. *Błędem średnim charakterystyki nazywa się średnie odchylenie dla odnośnej wtórnej populacji.* Otóż było podane w ustępie 28, że takie populacje zbliżają się szybko do formy normalnej przy powiększeniu zakresu prób. Normalny zaś rozkład częstości

ma tę właściwość, że blisko dwie trzecie częstości, dokładniej mówiąc część 0.6826 całości, mieści się w przedziale między ujemną a dodatnią wartością średniego odchylenia. Widoczne to jest z całki Laplace'a (zob. tabelę B podaną w końcu książki). Można zatem powiedzieć, że średni błąd danej charakterystyki określa w przybliżeniu granice, w których mieszczą się w dwóch przypadkach na trzy odchylenia tego rodzaju charakterystyk dla prób od ich średniej czyli od nadziei matematycznej danej charakterystyki. Widzieliśmy nadto w poprzednim ustępie, że nadzieja matematyczna jest zbliżona do wartości danej charakterystyki dla populacji, a w przypadku średniej arytmetycznej nawet jej równa. Wobec tego można powiedzieć, że średni błąd danej charakterystyki określa w przybliżeniu granice, w których mieszczą się w dwóch przypadkach na trzy odchylenia tego rodzaju charakterystyk dla prób od wartości danej charakterystyki dla populacji.

Średni błąd oznacza się grecką literą ε z podaniem symbolu odnośnej charakterystyki w nawiasach, np. średni błąd średniej arytmetycznej oznacza się symbolem ε *(m)*. Wartość średniego błędu dodaje się zwykle do wartości charakterystyki ze znakiem $\pm$, np. pisze się $m \pm \varepsilon(m)$.

Najprościej przedstawia się rozpatrywane zagadnienie dla średniej arytmetycznej; mianowicie można dowieść, że średnie odchylenie wtórnej populacji złożonej ze średnich dla prób równa się średniemu odchyleniu populacji pierwotnej podzielonemu przez pierwiastek kwadratowy z wielkości próby.

Istotnie widzieliśmy, że populacja 1, 2, 3, 4, 5 miała moment drugiego stopnia równy dwa, a więc średnie odchylenie $\sqrt{2}$. Obliczmy moment drugiego stopnia dla wtórnej populacji złożonej ze średnich dla trójkowych prób według danych tabeli

28,3. Otrzymamy wartość równą $\frac{2}{3}$, a więc średnie odchylenie równe $\sqrt{\frac{2}{3}}$ zgodnie z podanym powyżej twierdzeniem.

Będziemy zatem mieli dla obliczenia średniego błędu średniej arytmetycznej wzór $\varepsilon\,(m) = \frac{S}{\sqrt{n}}$, gdzie S jest średnim odchyleniem dla populacji, a n liczbą wariantów w próbach. W praktycznym wykonaniu obliczenia będzie się miało trudność, wynikającą stąd, że dokładne wartości charakterystyk populacji nie są znane. Otóż widzieliśmy w ustępie poprzednim, że wielkość $s = \sigma\sqrt{\frac{n}{n-1}}$ jest zbliżona do wielkości S. Wobec tego tę ostatnią wielkość używa się zamiast S przy obliczaniu średniego błędu średniej arytmetycznej, tak samo zresztą jak we wszystkich innych przypadkach, w których ma się do czynienia ze średnim odchyleniem dla populacji.

Zanim przejdziemy do rozpatrzenia średniego błędu innych charakterystyk, warto jest mimochodem rozpatrzyć stosowane czasem zamiast średniego błędu pojęcie błędu prawdopodobnego. Odznacza się on tym, że odchylenia wartości charakterystyki dla prób od jej nadziei matematycznej wypadają równie często większe jak i mniejsze od niego. Z całki Laplace'a wynika, że błąd prawdopodobny jest mniejszy od średniego w stosunku 0.67449... do 1. Użycie tej wielkości nie daje żadnych korzyści większych od użycia błędu średniego.

Równania średniego błędu dla innych charakterystyk są naogół bardziej złożone. I tak średni błąd momentu drugiego stopnia wyraża się wzorem:

$$\varepsilon\,(m_2) = \sqrt{\frac{M_4 - M_2^2}{n}}\,,$$

a średni błąd odchylenia średniego:

$$\varepsilon\,(\sigma) = S\sqrt{\frac{B-1}{4n}}\,,$$

gdzie B jest wskaźnikiem skupienia wariantów w populacji. Przy użyciu ekscesu zamiast wskaźnika skupienia wariantów wzór przybierze formę:

$$\varepsilon(\sigma) = S\sqrt{\frac{2-\Gamma_2}{4n}},$$

gdzie Γ_2 jest ekscesem dla populacji.

Dla populacji niezbyt odbiegających od formy normalnej eksces jest mały. Można go przyrównać zeru i wtedy wzór ulega uproszczeniu:

$$\varepsilon(\sigma) = \frac{S}{\sqrt{2n}}$$

Zagadnienie średniego błędu dla odchylenia przeciętnego jest bardziej zawiłe niż dla średniego. Wobec tego podam tylko przybliżony wzór dla populacji niezbyt silnie odbiegających od formy normalnej:

$$\varepsilon(\vartheta) = S\sqrt{\frac{\pi-2}{2n}}$$

Z tego wzoru wynika, że średni błąd dla przeciętnego odchylenia jest nieco większy niż dla średniego.

Dalej dla momentów trzeciego i czwartego stopnia będziemy mieli równania:

$$\varepsilon(m_3) = \sqrt{\frac{M_6 - M_3^2 - 6M_4M_2 + 9M_2^3}{n}}$$

$$\varepsilon(m_4) = \sqrt{\frac{M_8 - M_4^2 - 8M_5M_3 + 16M_2M_3^2}{n}}$$

Zawierają one momenty wyższych stopni, aż do ósmego. Oczywiście przy takiej złożoności nie mogą mieć znaczenia prak-

tycznego. Ale jeżeli populacja nie odbiega zbytnio od formy normalnej, następuje znaczne uproszczenie:

$$\varepsilon(m_3) = S^3 \sqrt{\frac{6}{n}}, \quad \varepsilon(m_4) = S^4 \sqrt{\frac{96}{n}}$$

Bardzo proste są wzory dla błędów wskaźnika skośności γ_1 i ekscesu γ_2:

$$\varepsilon(\gamma_1) = \sqrt{\frac{6}{n}} \quad \varepsilon(\gamma_2) = \sqrt{\frac{24}{n}} = 2\varepsilon(\gamma_1)$$

Różnią się one od wszystkich poprzednich tym, że żadne momenty w nich nie występują. Dla obliczenia ich wystarczy wiedzieć n — liczbę wariantów w próbie.

Zastosowanie wszystkich tych wzorów z wyjątkiem dwóch ostatnich jest utrudnione przez to, że w nich występują charakterystyki populacji. W praktycznym wykonaniu obliczeń, kiedy ma się tylko jedną próbę, niema innego wyjścia jak użycie charakterystyki próby zamiast charakterystyki populacji. Dla średniego odchylenia ta kwestia była dokładniej omówiona.

Zastosujmy tytułem przykładu powyższe wzory do liczby kwiatów w koszykach sałatnicy. Obliczmy najpierw średni błąd średniej arytmetycznej. Otóż mieliśmy dla tego materiału: $m = 17.28$, $\sigma = 2.603$, $n = 359$. Wypadnie $\varepsilon(m) = 0.14$ i $m = 17.28 \pm 0.14$.

Obliczmy dalej średni błąd średniego odchylenia. Ponieważ eksces $\gamma_2 = -0.315$, otrzymamy $\varepsilon(\sigma) = 0.107$ i $\sigma = 2.603 \pm 0.107$.

Wreszcie średnie błędy wskaźnika skośności i ekscesu będą

$$\varepsilon(\gamma_1) = 0.129 \qquad \varepsilon(\gamma_2) = 0.258$$

a zatem $\gamma_1 = -1.461 \pm 0.129$, $\gamma_2 = -0.315 \pm 0.258$.

Podobnie otrzymamy dla liczby kwiatów w szczytowych koszykach bławatka z Zimnej Wody:

$$m = 34.77 \pm 0.34$$
$$\sigma = 5.891 \pm 0.140$$
$$\gamma_1 = +0.731 \pm 0.141$$
$$\gamma_2 = +1.427 \pm 0.283$$

Na ostatku trzeba jeszcze rozpatrzyć zagadnienie średniego błędu dla różnicy między wartościami tej samej charakterystyki obliczonymi dla różnych materiałów statystycznych. Jest to zagadnienie bardzo zawiłe. Ograniczę się wobec tego do przedstawienia go odnośnie do średniej arytmetycznej. Niech będą dwie próby: jedna złożona z n_1 wariantów, druga złożona z n_2 wariantów. Te próby mogą pochodzić z tej samej populacji albo z różnych. Średnia dla pierwszej z nich niech będzie m', dla drugiej m''. Sumę kwadratów odchyleń od średniej dla pierwszej próby oznaczymy przez Σ_1, dla drugiej przez Σ_2. Wtedy średni błąd różnicy średnich $m' - m''$ wyrazi się wzorem

$$\varepsilon(m' - m'') = \sqrt{\frac{\dfrac{\Sigma_1 + \Sigma_2}{n_1 n_2 (n_1 + n_2 - 2)}}{n_1 + n_2}}$$

Jeżeli obie próby są jednakowej wielkości, to znaczy, że $n_1 = n_2$, powyższy wzór ulega uproszczeniu:

$$\varepsilon(m' - m'') = \sqrt{\frac{\Sigma_1 + \Sigma_2}{n(n-1)}}$$

W tym przypadku średni błąd różnicy średnich arytmetycznych równa się pierwiastkowi kwadratowemu z sumy kwadratów błędów dla średnich obu prób:

$$\varepsilon(m' - m'') = \sqrt{[\varepsilon(m')]^2 + [\varepsilon(m'')]^2}$$

Zastosujmy te wzory tytułem przykładu do podanych w ust. 6 materiałów dotyczącyh liczby kwiatów w szczytowych koszykach bławatka pochodzącego z dwóch różnych miejscowości. Potrzebne do tego sumy kwadratów odchyleń obliczymy według wzoru:

$$\Sigma (u_i - \bar{u})^2 = \Sigma \; u_i - a^2 - \frac{[\Sigma (u_i - a)]^2}{n}$$

w którym a jest wartością wyjściową (por. ust. 13).

Dla materiału z Zimnej Wody mamy $m' = 34.77$, $n_1 = 300$, $\Sigma (u_i - a)^2 = 10589$, $\Sigma (u_i - a) = 231$ przy $a = 34$.

Dla materiału z Dublan mamy: $m'' = 31.06$, $n_2 = 200$, $\Sigma (u_i - a)^2 = 5094$, $\Sigma (u_i - a) = 130$ przy $a = 31$.

W wyniku obliczeń otrzymamy: $\varepsilon (m' - m'') = \pm 0.51$ i $m' - m'' = 3.71 \pm 0.51$. Wypadła różnica między średnimi z górą 7 razy większa od swojego błędu. Tak wielką w porównaniu do swojego błędu różnicę można uważać za istotną. Można wogóle przyjąć za prawidło, że jeżeli różnica między wartościami charakterystyki dla dwóch prób przewyższa conajmniej trzykrotnie swój średni błąd, to można ją uznać za istotną, czyli nie pochodzącą z przypadku, lecz spowodowaną przez różnicę w charakterze materiałów, z których próby były wzięte. Jeżeli natomiast ta różnica jest mniejsza od trzykrotnego błędu, trzeba już liczyć się z możliwością przypadkowego zestawienia prób z tej samej populacji. Tak właśnie będzie, jeżeli porównamy ze sobą nie całości materiałów z Zimnej Wody i Dublan, lecz tylko koszyki „ósemkowe" — z ośmioma obwodowymi płonnymi kwiatami (tab. 30,1).

Tabela 30,1

Liczby kwiatów w szczytowych koszykach bławatka z 8 płonnymi obwodowymi kwiatami.

Liczby kwiatów	Ich częstości w materiale		Liczby kwiatów	Ich częstości w materiale	
	z Zimnej Wody	z Dublan		z Zimnej Wody	z Dublan
21	1	—	z przen.	53	33
22	1	—	32	11	5
23	1	—	33	9	4
24	—	1	34	6	5
25	2	1	35	4	4
26	1	1	36	4	2
27	3	4	37	—	1
28	13	10	38	2	3
29	10	9	39	2	—
30	9	2	42	—	1
31	12	5	45	—	1
do przen.	53	33	Ogółem	91	59

Mamy teraz dla materiału z Zimnej Wody $m' = 30.88$, $n_1 = 91$, $\Sigma (u_i - a)^2 = 1116$, $\Sigma (u_i - a) = 80$ przy $a = 30$. Dla materiału zaś Dublańskiego: $m'' = 31.37$, $n_2 = 59$, $\Sigma (u_i - a)^2 = 982$, $\Sigma (u_i - a) = 22$ przy $a = 31$. Z tych danych otrzymamy dla różnicy średnich arytmetycznych średni błąd $\varepsilon (m' - m'') = \pm 0.62$, a zatem $m' - m'' = -0.49 \pm 0.62$. Różnica średnich wypadła tu mniejsza od swojego błędu. By ocenić jej wartość, trzeba obliczyć prawdopodobieństwo pochodzenia jej z przypadku. To zagadnienie będzie omówione w następnym rozdziale.

31. Wiarogodność charakterystyk małych prób. Z wywodów poprzedniego ustępu wynika, że błąd średni tylko przy wielkich próbach rozwiązuje zagadnienie wiarogodności charakte-

rystyk. Dla prób małych, z jakimi na przykład wypada pracować w doświadczalnictwie rolniczym, trzeba szukać innej drogi. Nie można tu zastępować wartości charakterystyk populacji przez ich wartości dla próby. Trzeba częstości względne, inaczej mówiąc prawdopodobieństwa statystyczne różnic między charakterystykami próby i populacji wyznaczać bezpośrednio według właściwości próby. Rozwiązanie postawionego w ten sposób zagadnienia wypadnie naturalnie różnie, zależnie od charakteru populacji. W dotychczasowych pracach nad tym zagadnieniem przyjmowało się, że populacja ma charakter normalny. Na szczęście okazało się, że dla innego rodzaju populacji wyniki zachowują swoje znaczenie, o ile rozkład częstości różnych wartości wariantów w populacji nie zanadto odbiega od rozkładu normalnego.

Dla średniej arytmetycznej zagadnienie zostało rozwiązane przez Williama Gosseta (1876—1937) piszącego pod pseudonimem Student. Stopień zgodności między średnią próby a średnią populacji określił on za pomocą funkcji $t = \frac{m - M}{\varepsilon\, m}$, gdzie m jest średnią dla próby, M — średnią dla populacji, $\varepsilon(m)$ — średnim błędem średniej arytmetycznej próby. Wartości funkcji t mogą być dodatnie i ujemne stosownie do znaku różnicy $m - M$. Absolutna wartość ich jest na ogół tym mniejsza, im średnia próby jest bliższa średniej dla populacji, ale nie zawsze.

Na przykład jeżeli weźmiemy znowu populację 1, 2, 3, 4, 5, to na podstawie tabeli 28,2 otrzymamy dla 35 rodzajów prób następujące wartości t (tab. 33,1). W tej tabeli nieokreśloną wartość $\frac{0}{0}$ dla prób 333 przyjmiemy równą zeru. Z niej widzimy na przykład, że próby 334 dają średnią różniącą się od średniej populacji o ⅓, próby zaś 155 średnią różniącą się o ⅔ a jednocześnie t dla pierwszego rodzaju prób jest równe 1.0, podczas gdy dla drugiego rodzaju ma wartość 0.5 — nie większą zatem, jak by się należało spodziewać, lecz mniejszą.

Tabela 31,1

Rodzaje prób	Wartości $m - M$ dla tych prób w $^1/_3$	t
135	0	0
144		
225		
234		
333		
145	+1	0.2773
235		0.3780
244		0.5000
334		1.0000
155	+2	0.5000
245		0.7560
335		1.0000
344		2.0000
255	+3	1.0000
345		1.7321
444		∞
355	+4	2.0000
445		4.0000
455	+5	—5.0000
555	+6	∞
125	—1	—0.2773
134		—0.3780
224		—0.5000
233		—1.0000
115	—2	—0.5000
124		—0.7560
133		—1.0000
223		—2.0000
114	—3	—1.0000
123		—1.7321
222		∞
113	—4	—2.0000
122		—4.0000
112	—5	—5.0000
111	—6	∞

Wobec powyższego w dalszym ciągu tych wywodów będziemy nazywać „lepszymi" próby mające t o wartościach absolutnych mniejszych, jakkolwiek ich średnie nie zawsze będą bliższe średniej dla populacji niż średnie prób „gorszych".

W tym założeniu S t u d e n t dowiódł, że częstość względna występowania wartości funkcji t w zespole prób wziętych z populacji o formie normalnej wyraża się funkcją:

$$s(t) = C\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}.$$

w której współczynnik C zależy tylko od wielkości próby. Można go obliczyć za pomocą funkcji „gamma", oznaczanej symbolem $F(x)$:

$$C = \frac{\left(\Gamma\frac{n}{2}\right)}{\sqrt{\pi(n-1)\Gamma\left(\frac{n-1}{2}\right)}}$$

Funkcja gamma określa się równaniem

$$\Gamma\, x+1 = \lim_{n=\infty} \frac{1\cdot2\cdot3\ldots(n-1)n}{x(+1)(x+2)\ldots(x+n)}\, n^x$$

Funkcja $s(t)$ daje jednakowe wartości dla dodatnich i ujemnych wartości zmiennej niezależnej, podobnie jak funkcja de Moivre'a. Tak samo jak ta ostania, funkcja $s(t)$ maleje nieograniczenie ze zwiększeniem absolutnej wartości t. Ze zwiększeniem wielkości próby zbliża się ona do funkcji de Moivre'a i dla $n=\infty$ przybiera znaną już nam formę:

$$s(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Za pomocą rachunku całkowego można obliczyć dla każdej wartości t częstość względną prób o wartościach tej zmiennej mniejszych od danej; przy czym naturalnie wszystkie wartości ujemne trzeba uważać za mniejsze od najmniejszych nawet dodatnich. Uskutecznia się to za pomocą całki

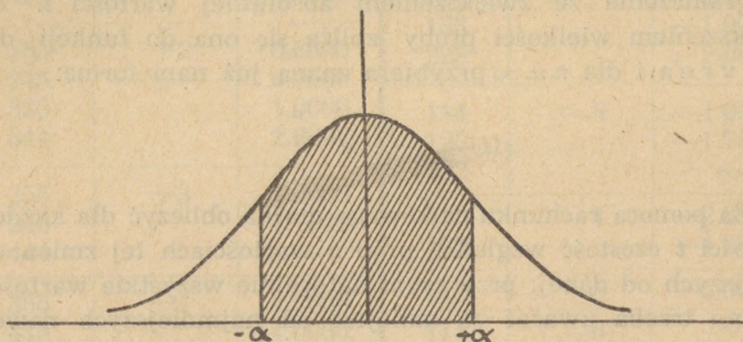
$$S(t) = C\int_{-\infty}^{t} \left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt$$

analogicznej całki Laplace'a. Tak samo jak ta ostatnia, całka $S(t)$ określa pole figury, zakreślonej przez oś odciętych i krzywą funkcji $s(t)$, licząc od lewego jej końca do rzędnej odpowiadającej danej wartości t (por. ryc. 5).

Tak samo jak dla całki Laplace'a, pole całej tej figury równa się jedności, a rzędna $t=0$ dzieli to pole na dwie równe części. W ten sposób $S(-\infty)=0$, $S(0)=0{,}5$, $S(+\infty)=1$.

Wartość całki $S(t)$ dla $t \geqq 0$ są podane w końcu książki w tabeli *E*. Przy korzystaniu z tej tabeli przeważnie jest obojętne czy średnia próby jest większa, czy mniejsza od średniej dla populacji, chodzi tylko o to, jak bliska jest tej ostatniej. Stąd wypada opierać się na absolutnej wartości t bez uwzględnienia znaku i poszukiwać częstość względną prób, dla których wartość bezwzględna t jest mniejsza od wartości bezwzględnej tej funkcji dla danej próby. Dlatego też w tabeli *E* są pominięte wartości funkcji $S(t)$ dla wartości $t < 0$.

Niech będzie a absolutną wartością funkcji t dla danej próby. Wtedy częstość względna prób, mających t o wartości absolutnej mniejszej od a, czyli pole prób „lepszych" od danej, będzie równa polu wycinka figury krzywej, ograniczonej przez



Ryc. 8.

rzędne $t = -a$ i $t = +a$ (ryc. 8). Ta częstość wyraża się wzorem: $P\{|t| < a\} = 2\,S(a) - 1$, w którym literą P jest oznaczone prawdopodobieństwo zależności przedstawionych przez symbole zawarte w nawiasach $\{\ \}$. Będzie to prawdopodobieństwo statystyczne prób lepszych od danej.

Na podstawie powyższego wzoru można wyznaczyć takie a, żeby próby mające t zawarte w granicach od $-a$ do $+a$ stanowiły z góry określoną część ogółu prób. Wielkość tej

części może być obrana dowolnie i nosi nazwę współczynnika ufności. Jeżeli oznaczymy ten współczynnik przez β ($\beta < 1$), to będziemy mieli równanie $\beta = 2\,S(\alpha) - 1$. Z niego wyprowadzamy zależność $S(\alpha) = \frac{1+\beta}{2}$, która da możność wyznaczania α według tabeli dla funkcji $S(t)$.

Dalej mamy $|t| = \frac{|m - M|}{\varepsilon(m)}$. Wobec tego będziemy mieli $\frac{|m - M|}{\varepsilon(m)} = \alpha$. Stąd wypływa równanie

$$P\{|m - M| < \alpha\,\varepsilon(m)\} = \beta,$$

z którego wynika, że nieznana średnia populacji mieści się w granicach między $m - \alpha\,\varepsilon(m)$ i $m + \alpha\,\varepsilon(m)$ z prawdopodobieństwem równym współczynnikowi ufności. W praktyce współczynnik ufności przyjmuje się zwykle równy 0.95 albo 0.99. Dla tej drugiej wartości wypada α większe i przeto granice zakreślone dla średniej populacji szersze. Wybór takiego czy innego współczynnika ufności zależy od ważności zadania oraz od psychologii badacza. Dla ważniejszych zagadnień bierze się współczynnik większy. Ostrożniejsi badacze wolą większe współczynniki dla większej pewności — ostrożność nigdy nie zawadzi!

Zastosujmy tę metodę do konkretnego przykładu. Otóż w pewnym doświadczeniu wykonanym na torfowisku Czemerne pod Sarnami w r. 1931 [1]) sucha masa trawy na 4 poletkach z pełnym nawożeniem o powierzchni 50 m^2 wyniosła w kilogramach 55.1, 50.3, 58.4, 45.6. Wypada z tych danych średnia arymetyczna 52.35, jej średni błąd 2.78. Wynikało by z tego, że gdyby doświadczenie było wykonane nie na czterech poletkach, ale na bardzo wielkiej ich liczbie i gdybyśmy wzięli z takiej populacji nie jedno doświadczenie, 4-poletkowe, lecz wielką ich ilość, to dwie trzecie tych doświadczeń dałoby śred-

---

[1]) D. Szymkiewicz i B. Świętochowski. Doświadczenia nad żyznością torfów. — Rocznik łąkowy i torfowy. 1936.

nie plony zawarte w granicach od 52.35—2.78 do 52.35 + 2.78. Tak wypadło by, gdyby średni błąd był dokładnie obliczony. Otóż było już wyjaśnione w ust. 30, że to tak nie jest, jeżeli próba jest mała. Za pomocą funkcji $S(t)$ możemy dokładniej obliczyć wspomniane granice.

W tym celu trzeba przyjąć dla współczynnika ufności wartość ⅔. Otrzymamy z równania $S(\alpha) = \frac{1+\beta}{2}$ dla $S(\alpha)$ wartość 0.833. Z tabeli funkcji $S(t)$ przez interpolację wyprowadzamy wartość dla α równą 1.15 i wobec tego granice, w których mieszczą się średnie dwóch trzecich prób będą szersze w stosunku 1.15 : 1, a więc wyniosą 52.35 — 3.20 i 52.35 + 3.20.

Ale współczynnik ufności ⅔ jest za mały. W praktyce bierze się conajmniej 0.95. Wtedy wypada $S(\alpha) = 0.975$, $\alpha = 3.2$ i granice, w których mieszczą się średnie 95% prób, będą 52.35 — 8.90 i 52.35 + 8.90. W tych też granicach z prawdopodobieństwem 0.95 będzie się mieściła średnia całej populacji złożonej z wyników doświadczeń na nieograniczonej liczbie poletek. Prawdopodobieństwo 0.95 oznacza, że 5% prób 4-poletkowych dałoby inny wynik — inne granice dla średniej arytmetycznej populacji. Nigdy nie jest wykluczone, że się na taką próbę natrafi — jak się ma pecha, to nic na to się nie poradzi.

Dla ułatwienia omawianych obliczeń podaję za Romanowskim tabelę 31,2 wartości α dla dwu najczęściej używanych współczynników ufności 0.95 i 0.99 przy różnym zakresie prób, począwszy od prób czwórkowych. Dla prób o zakresie większym od 20 można używać wartości dla prób nieskończonej wielkości.

Funkcja $s(t)$ i wywodząca się z niej $S(t)$ znajdują jeszcze dalsze i to ważniejsze zastosowanie. Mianowicie pozwalają one stwierdzić z określonym prawdopodobieństwem, czy różnica między średnimi arytmetycznymi dwóch prób pochodzi z przypadku, czy też stąd, że próby były wybrane z różnych populacji.

Będziemy rozumowali w sposób następujący. Wybieramy z populacji wszelkie możliwe próby po n_1 wariantów, następnie wszelkie możliwe próby po n_2 wariantów. Każdą próbę pierwszej grupy zestawiamy kolejno ze wszystkim próbami drugiej. Dla każdej pary prób obliczamy wartość funkcji

$$t' = \frac{m' - m''}{\varepsilon(m' - m'')}$$

gdzie m' jest średnią arytmetyczną jednej próby, m'' taka sama charakterystyka drugiej próby a $\varepsilon(m' - m'')$ — średni błąd różnicy średnich, obliczony według wzoru podanego w ust. 30.

Tabela 31,2

Wartości a dla różnych wielkości prób (n)

n	Współczynnik ufności		n	Współczynnik ufności	
	0.95	0.99		0.95	0.99
4	3.18	5.85	14	2.16	3.01
5	2.78	4.60	15	2.15	2.98
6	2.57	4.04	16	2.13	2.95
7	2.45	3.70	17	2.12	2.92
8	2.37	3.50	18	2.11	2.90
9	2.31	3.36	19	2.10	2.89
10	2.26	3.25	20	2.09	2.87
11	2.23	3.17	21	2.09	2.86
12	2.20	3.11	∞	1.96	2.58
13	2.18	3.06			

Widoczne jest od razu, że funkcja t' jest podobna do rozpatrzonej poprzednio funkcji

$$t = \frac{m - M}{\varepsilon(m)}$$

Może nawet być z tej ostatniej wyprowadzona. Istotnie jeżeli weźmiemy populację złożoną z różnic $m' - m''$, to średnia arytmetyczna takiej populacji będzie równa zeru, gdyż dla każdej różnicy $m' - m''$ można dobrać różnicę $m'' - m'$ równą, ale ze znakiem odwrotnym. Można więc założyć we wzorze funkcji t

$M = O$. Wystarczy teraz zamiast średnich m wprowadzić różnice średnich $m' - m''$, by z równania funkcji t otrzymać równanie dla funkcji t'.

Wartość absolutna $|t'|$ jest na ogół tym mniejsza, im mniejsza jest wartość absolutna $|m' - m''|$. Zależy ona jednak także od wielkości błędu $\varepsilon(m' - m'')$. Jeżeli przeto będziemy porównywali różne pary prób, może się zdarzyć, że jedna para prób będzie miała większe $|t'|$ a mniejsze $|m' - m''|$ niż druga para. Pod tym względem występują tu podobne zdarzenia jak dla funkcji t, gdzie dla większych $|t|$ trafiają się mniejsze różnice $|m - M|$. Można powiedzieć, że dwie próby są tym bardziej „zgodne", im mniejsze jest dla nich $|t'|$, a to w tym danym sensie, w jakim nazywaliśmy „lepszymi" próby o mniejszym $|t|$.

Przy powyższych założeniach można dowieść, że częstości względne wartości t' wyrażają się tą samą funkcją Studenta, z jaką mieliśmy do czynienia poprzednio. Będziemy tedy mieli funkcję:

$$s(t) = C\left(1 + \frac{t'^2}{n-1}\right)^{-\frac{n}{2}}$$

w której $n = n_1 + n_2 - 1$. Będziemy także mieli funkcję:

$$S(t') = C\int_{-\infty}^{t'}\left(1 + \frac{t'^2}{n-1}\right)^{-\frac{n}{2}} dt',$$

która będzie dawała częstości względne dla wartości t' mniejszych od danej.

Jeżeli teraz mamy jakąś parę prób o wartości t', której absolutna wielkość równa się a, to częstość względna par prób bardziej od niej „zgodnych" wyrazi się równaniem $P\{|t'| < a\} = 2S(a) - 1$, takim samym jakie mieliśmy dla częstości prób „lepszych" niż ta próba, dla której $|t| = a$.

Po tych wstępnych wywodach powróćmy do naszego zagadnienia. Mamy parę prób. Jeżeli one pochodzą z tej samej populacji, to nie powinno być zbyt dużo par prób zgodniejszych od nich, bo takie próby odznaczają się większą zgodnością niż

próby wzięte z różnych populacyj. Można wobec tego wybrać pewną granicę, której częstość prób zgodniejszych niż dane nie powinna przekraczać. Można ją nazwać tak samo współczynnikiem ufności, jak poprzednio przy rozpatrywaniu odchyleń średnich dla prób od średniej dla populacji. Jeżeli ta granica nie jest przekroczona, można będzie uznać, że próby pochodzą z tej samej populacji. Jeżeli zaś częstość prób zgodniejszych niż dane wypadnie większa od współczynnika ufności, to trzeba będzie uznać, że próby pochodzą z różnych populacji. Trzeba zatem dla danej pary prób obliczyć wartość *t'* i według tabeli funkcji $S(t)$ obliczyć prawdopodobieństwo bardziej zgodnych prób według przytoczonego poprzednio równania $P\{|t'|<\alpha\} = 2S(\alpha) - 1$. Badanie sprowadza się ostatecznie do sprawdzenia czy prawdopodobieństwo statystyczne $P\{|t'|<a\}$ jest mniejsze. czy większe od współczynnika ufności, który wybiera się zwykle w wymiarze 0.95 albo 0.99, tak samo jak poprzednio.

Zastosujmy powyższe prawidło do konkretnych przykładów. I tak we wspomnianych powyżej kulturach na torfowisku Czemernem oprócz pełnego potasowo-azotowo-fosforowego nawożenia |*KNP*| stosowano także nawożenie potasowo-azotowe bez fosforu |*KN*| i azotowo-fosforowe bez potasu |*NP*|. Wszystkie doświadczenia były 4-poletkowe. Nawożenie |*KN*| dało plony siana 47.8, 47.6, 46.0, 47.2 kg, nawożenie |*NP*| 29.9, 19.0, 22.2, 24.7. Pierwsze dało średnią 47.15, ± 0.40, drugie 23.95 ± 2.30. Zestawiając te dane z plonami otrzymanymi na pełnym nawożeniu, otrzymamy różnice:

$$|KNP| - |KN| = 5.20 \pm 2.81$$
$$|KNP| - |NP| = 28.40 \pm 3.61$$

Stąd dla pary prób [*KNP*] — [*KN*] otrzymamy $t' = 1.85$, dla pary zaś |*KNP*| — [*NP*] $t' = 7.87$. Ponieważ zakres prób obejmował 4 warianty, mamy $n = 7$. Stąd według tabeli funkcji $S(t)$ znajdziemy dla rozpatrywanych par prób prawdopodobieństwa statystyczne

$$P\{|t'| < 1.85\} = 0.886$$
$$P\{|t'| < 7.87\} = 1.000$$

Wnosimy stąd, że próby pierwszej pary można uważać za pochodzące z tej samej populacji, próby zaś drugiej za pochodzące z różnych populacyj. W ten sposób różnica w plonach wynikająca z niezastosowania nawozów fosforowych okazała się przypadkową, natomiast różnica powodowana przez brak potasu była istotna.

Omawiana metoda porównywania średnich dla prób może być stosowana także do dużych prób. Zastosujmy ją do liczb kwiatów w koszykach bławatka z Zimnej Wody i Dublan (tab. 5 (1) oraz 5,1 i 30,1). Tu według obliczeń ust. 30 dla całości obu materiałów wypada $m' - m'' = 3.71 \pm 0.51$. Stąd $|t'| = 7.27$. Wypada P prawie równe jedności. Te materiały możemy uznać za pochodzące z różnych populacyj nawet przy największym współczynniku ufności. Natomiast dla koszyków z ośmioma płonnymi kwiatami obwodowymi otrzymaliśmy $m' - m'' = -0.49 \pm 0.62$ a zatem $|t'| = 0.79$ *i* $P = 0.785$. Można więc przyjąć, zarówno przy współczynniku ufności 0.95, jak i 0.99, że „ósemkowe" koszyki pochodzą z tej samej populacji.

Podobna do powyższej metoda oceny wiarogodności średniej arytmetycznej została przez Studenta i Fishera zastosowana do średniego odchylenia. Rolę funkcji t odgrywa tu inna funkcja. Nie będziemy rozpatrywali tego zagadnienia, gdyż znaczenie jego jest niewielkie. Odnośne informacje znajdzie czytelnik w podręcznikach Fishera i Romanowskiego.

32. Krańcowe warianty. Trzeba omówić jeszcze ostatnią kwestię dotyczącą wiarogodności charakterystyk. Wartości niektórych z nich zależą w silnym stopniu od krańcowych wariantów, mianowicie wartości momentów wyższych rzędów. Pochodzi to stąd, że częstości krańcowych wariantów przy obliczaniu takich momentów są mnożone przez wysokie potęgi odchyleń. Stąd już zmiana tych częstości o jedynkę wystarcza, by w wyniku wystąpiła poważna różnica, a takie zmiany skutkiem przypadkowości w doborze wariantów są na porządku dziennym.

Szczególnie duże różnice w wartościach momentów występują wtedy, kiedy materiał jest nieliczny. Jako przykład tego rodzaju zjawisk można przytoczyć odosobniony pojedyńczy wariant o wartości 62 w materiale bławatka. Jeżeli go odrzucimy, to przy trzecim sposobie grupowania otrzymamy: $m' = 11.547$ zamiast 11.576, $m'_2 = 3.630$ zamiast 3.958, $m'_3 = +3.519$ zamiast $+5.424$ i $m'_4 = 45.42$ zamiast 64.35.

Dla zaradzenia złu, powodowanemu przez niepewności w liczbie krańcowych wariantów, trzeba je odrzucać, ale należy postępować ostrożnie, kierując się pewnym prawidłem. Chodzi tu o pojedyńcze warianty wysokiej względnie niskiej wartości, takie jak przytoczony powyżej wariant 62 w materiale bławatka. Prawdopodobieństwo występowania takich wariantów jest małe. Przeto wystąpią one tylko w części prób wybranych z danej populacji. Jeżeli jednak skutkiem przypadkowego zbiegu okoliczności wystąpią w rozpatrywanej próbie, można je odrzucić, jeżeli ich średnia częstość w próbach danej wielkości jest mniejsza niż ½, gdyż wtedy w większości prób one nie ukażą się wcale. Jest to prawidło zaproponowane przez Chauvenet'a. Trzeba je stosować osobno dla krańcowych plus wariantów i minus wariantów.

Dla zastosowania prawidła Chauvenet'a trzeba znać ogólny rozkład częstości w rozpatrywanym materiale. Dokładna znajomość tego rozkładu jest oczywiście niemożliwością. Ale wystarczy znajomość ogólnikowa. W szczególności łatwo jest stosować prawidło Chauvenet'a do szeregów rozdzielczych o formie zbliżonej do normalnej. Wtedy za pomocą całki Laplace'a można obliczyć prawdopodobieństwo występowania skrajnych wariantów, a mnożąc to prawdopodobieństwo przez liczbę wariantów w rozpatrywanej próbie, można znaleźć średnią częstość takich wariantów w próbach danej wielkości. Oznaczając przez p wspomniane prawdopodobieństwo, przez n liczbę wariantów w próbie, otrzymujemy równanie $np = {}^1/_2$, z którego wypływa $p = \frac{1}{2n}$. Według tej wartości p z tabeli całki Laplace'a znajdujemy wartość zmiennej ξ,

której dane p odpowiada. Jak to już nam jest wiadomo z ust. 25, zmienna ξ jest związana ze zmienną ewentualną równaniem $\xi = \frac{x-m}{\sigma}$, z której wypływa $x = \sigma\xi + m$. Jeżeli obliczona w ten sposób wartość x wynosi a, to trzeba odrzucić minus warianty o wartości $< a$ i plus warianty o wartości większej niż $a + m$.

Dla przeprowadzenia takich obliczeń potrzebne są wartości całki Laplace'a dla wysokich wartości zmiennej ξ z 5 znakami dziesiętnymi. Tabela B do tego nie zawsze wystarcza i dlatego podaję dodatkową tabelę 32,1.

Zastosujmy powyższe wywody do materiału bławatka zgrupowanego trzecim sposobem. Ponieważ liczba wariantów jest tu 300, wypada $p = 1/600 = 0.00167$. Z tabeli 32,1 wynika, że $\xi = 2.93$, a ponieważ $m' = 11.576$ i $\sigma' = 1.968$, otrzymujemy $= 5.766$. Trzeba zatem odrzucić minus warianty o wartości x' mniejszej niż 5.766 i plus warianty o wartości x' większej niż 17.342. Wypadnie więc zachować wszystkie minus warianty i odrzucić krańcowy plus wariant 62 z klasy $x' = 20\,1/3$, o którym była mowa poprzednio.

Tabela 32,1

Wartość całki Laplace'a w stutysięcznych

ξ	0	1	2	3	4	5	6	7	8	9
–2.9	187	181	175	169	164	159	154	149	144	139
–3.0	135	131	126	122	118	114	111	107	103	100
–3.1	97	94	90	87	84	82	79	76	74	71
–3.2	69	66	64	62	60	58	56	54	52	50
–3.3	48	47	45	43	42	40	39	38	36	35
–3.4	34	32	31	30	29	28	27	26	25	24
–3.5	23	22	22	21	20	19	19	18	17	17
–3.6	16	15	15	14	14	13	13	12	12	11
–3.7	11	10	10	10	9	9	8	8	8	8
–3.8	7	7	7	6	6	6	6	5	5	5
–3.9	5	5	4	4	4	4	4	4	3	3

Tabela 32,2

Częstości (w tysięcznych) temperatur minimalnych powietrza w trzech dekadach maja

Temperatury	Kraków			Tarnopol			Wilno		
	1	2	3	1	2	3	1	2	3
(–4.9)–(–0.1)	20	5	—	10	10	—	66	43	16
0.0 — 4.9	175	130	41	265	115	68	359	325	97
5.0 — 9.9	670	605	350	425	425	336	407	380	378
10.0 — 14.9	135	255	554	285	390	564	156	252	449
15.0 — 19.9	—	5	55	15	60	32	12	—	60
Średnie temperatur	7	8.6	10.7	7.4	9.2	10.3	5.9	6.6	9.7

Na tym jeszcze nie koniec. Bywają przypadki, w których nie ma mowy o odrzucaniu skrajnych wariantów, kiedy przeciwnie — mają te warianty największe znaczenie. Tak jest na przykład z minimalnymi temperaturami powietrza w wiosennych miesiącach, kiedy częstości krańcowych najniższych temperatur schodzących poniżej zera określają częstość zgubnych dla roślinności wiosennych przymrozków. Ilustracją może być tabela 32,2 obliczona dla Krakowa, Tarnopola i Wilna na podstawie 20-letnich obserwacyj.

ROZDZIAŁ VI.

WYRÓWNANIE SZEREGÓW ROZDZIELCZYCH.

33. Ogólne zasady. Widzieliśmy na przykładzie bławatka w ust. 5, że ciągi częstości w szeregach rozdzielczych wykazują zazwyczaj przeskoki, powodowane przez przypadkowość prób. Dla wyrównania tych przeskoków łączy się wartości zmiennej ewentualnej w klasy, których szerokość wybiera się o tyle dużą, ażeby częstość zmieniała się stopniowo. Pociąga to tę niedogodność, że zatracają się częstości dla poszczególnych wartości zmiennej ewentualnej. Tymczasem są one nieraz potrzebne dla różnych celów, zwłaszcza praktycznych, np. w ubezpieczeniach. Są m. inn. potrzebne dla wyznaczenia wartości modalnej. Dla zaradzenia złemu trzeba przedstawić szeregi w formie funkcyj, które dadzą możność obliczenia częstości dla każdej wartości zmiennej ewentualnej, np. w przypadku bławatka częstości poszczególnych liczb kwiatów w koszykach.

Poza tym tego rodzaju funkcje charakteryzują zmienność materiałów statystycznych i mogą zastąpić omówione w rozdziałach II i III charakterystyki.

Jakie funkcje mogą być użyte do zobrazowania szeregów rozdzielczych? Pierwszy warunek po temu jest, żeby nie przybierały wartości ujemnych, bo częstości są zawsze dodatnie. Poza tym w zasadzie każda funkcja może być użyta. Trzeba jeszcze dobrać taką, któraby dostatecznie dokładnie przedstawiała szereg rozdzielczy. Nie chodzi przy tym bynajmniej o to, by dokładność była zupełna, co jest łatwo osiągnąć za pomocą funkcji L a g r a n g e'a. Trzeba bowiem mieć na uwadze, że ma się zwykle do czynienia z przypadkowymi próbami i przeto stojące do dyspozycji wartości częstości nie są dokładne. Należy zatem wybrać takie funkcje, któreby wyrównywały przy-

padkowe odchylenia częstości w próbie do odnośnych częstości w populacji. Stąd wynika, że przy wyborze funkcyj trzeba kierować się nauką o prawdopodobieństwie.

Najgłębsze badania nad omawianym zagadnieniem przeprowadził Charlier[1]), opierając się na zasadach wyłożonych w wiekopomnym dziele Laplace'a „Théorie analitique des probabilités". W tych badaniach traktował Charlier specjalnie kwestię błędów obserwacyjnych, ale jego sposób rozumowania może być zastosowany także do innych podobnych faktów. Można mianowicie rozumować w sposób następujący. Mamy taką czy inną zmienną — mogą to być wyniki pomiarów, liczby kwiatów itp. Wartości tej zmiennej zależą z jednej strony od przyczyn działających stale, z drugiej zaś strony od przyczyn występujących od czasu do czasu. Mamy tu te same dwa rodzaje przyczyn, z którymi zaznajomiliśmy się w nauce o prawdopodobieństwie i które były nazwane określonymi i nieokreślonymi. Te ostatnie mogą być także nazwane przypadkowymi. Gdyby przyczyn przypadkowych nie było, zmienna miałaby stałą niezmienną wartość. Działania tych przyczyn powodują odchylenia od wspomnianej stałej wartości. Nie znając bliżej przyczyn przypadkowych, można wielkość i częstość odchyleń przedstawić za pomocą pewnych funkcyj, jeżeli się założy, że przyczyny przypadkowe są liczne i że działanie każdej z nich z osobna jest słabe. Przy takich założeniach Charlier dowiódł, że zmienność powodowana przez przyczyny przypadkowe da się zobrazować za pomocą dwóch rodzajów funkcyj, nazwanych przez niego „typem A" i „typem B".

Pierwszy z tych typów wyprowadza się z funkcji de Moivre'a i charakteryzuje szeregi dwuboczne niezbyt odbiegające od formy normalnej. Sama funkcja de Moivre'a stanowi poszczególny przypadek tego typu. Typ B wyprowadza się z szeregu Poissona i stosuje się do szeregów jedno-

[1]) „Ueber das Fehlergesetz" (Meddelanden från Lunds Astronomiska Observatorium N. 25) oraz „Die zweite Form des Fehlergesetzes" (ta sama seria praw N. 26).

bocznych oraz do dwubocznych silnie odbiegających od formy normalnej. Naturalnie sam szereg Poissona należy do tego typu. Typ *B* stopniowo przechodzi w typ *A* przy zwiększeniu charakterystycznej dla niego stałej λ.

Wszystkie te funkcje wywodzą się, jak to było wyjaśnione w ust. 25 i 26, z szeregu dwumianowego. W pewnych przypadkach sam ten szereg może służyć do wyrównywania szeregów rozdzielczych.

W praktycznym zastosowaniu po wyborze rodzaju funkcji trzeba wybrać określoną jej formę. Do tego służy oczywiste prawidło, że średnia arytmetyczna oraz momenty funkcji służącej do wyrównywania szeregu rozdzielczego powinny być równe odnośnym charakterystykom szeregu.

Dodatkowo należy zaznaczyć, że liczni statystycy proponowali wiele różnych funkcji do wyrównywania szeregów rozdzielczych. Najdokładniej tę kwestię rozważał Pearson. Zaproponował on użycie oprócz szeregu dwumianowego i funkcji de Moivre'a jeszcze siedmiu innych funkcji. Obszerny wykład tych funkcyj znajdzie czytelnik w książce W. Palin Eldertona „Frequency curves and correlation" (3 wyd. Cambridge 1838).

34. Szereg dwumianowy. Rozpoczniemy od szeregu dwumianowego, z którego wywodzą się, jak widzieliśmy, inne funkcje charakteryzujące zmienność szeregów rozdzielczych. Pomimo swojego podstawowego charakteru może on być stosowany tylko w bardzo ograniczonym zakresie. Przyczyną tego jest mała rozmaitość jego form. Jak to było wyjaśnione w ust. 24, formy te są określone przez dwie tylko wielkości n i p. Tymczasem dla charakterystyki materiałów statystycznych trzeba użyć co najmniej czterech: a mianowicie średniej arytmetycznej oraz momentów drugiego, trzeciego i czwartego stopnia. Średnie arytmetyczne można wyrównać przez odpowiednią zamianę zmiennej ewentualnej, mamy więc do obliczenia dwóch charakterystycznych stałych szeregu dwumianowego trzy równania.

Oczywiście wielkości otrzymane z dwóch równań nie muszą czynić zadość trzeciemu. Jeżeli naprzykład szereg dwumianowy będzie miał właściwe momenty drugiego i trzeciego stopnia, to na ogół wykaże inną wartość dla momentu czwartego stopnia niż rozpatrywany szereg rozdzielczy.

Właściwe swoje zastosowanie znajduje szereg dwumianowy w zagadnieniach dotyczących powtarzania się zdarzeń, z których to zagadnień został wyprowadzony. Rozpatrzymy w tym względzie dwa przykłady zaczerpnięte z książki Fishera.

Pierwszy przykład dotyczy gry w kości. Rzucano naraz 12 kości i notowano, na ilu kościach ukazywały się liczby 5 i 6. Rzutów dokonano 26 306. Segregując te rzuty według liczby kości z numerami 5 albo 6, otrzymano interesujący szereg rozdzielczy przedstawiony w tabeli 34,1 wraz z różnymi liczbami, odnoszącymi się do danego zagadnienia.

Tabela 34,1

Rzuty dwunastoma kośćmi

Rzuty o różnych ilościach kości z numerami 5 albo 6	Liczby takich rzutów		Różnice	Liczby rzutów obliczone dla wykrzywionych kości	Różnice względem liczb obserwowanych
	obserwowane	obliczone dla kości prawidłowych			
0	185	203	−18	187	−2
1	1149	1216	−67	1146	−3
2	3265	3345	−80	3215	+50
3	5475	5576	−101	5465	+10
4	6114	6273	−159	6269	−155
5	5194	5018	+176	5115	+76
6	3067	2927	+140	3043	+24
7	1331	1255	+76	1330	+1
8	403	392	+11	424	−21
9	105	87	+18	96	+9
10	14	13	+1	15	−1
11	4	1	+3	1	+3
12	0	0	0	0	0
Ogółem	26306	26306	−425 +425	26306	−179 +179

W trzeciej kolumnie tej tabeli są podane liczby rzutów obliczone teoretycznie za pomocą szeregu dwumianowego. Chodzi tu mianowicie o powtarzanie się pewnego zdarzenia — wystąpienia numeru 5 albo 6 — w dwunastkowych zbiorach przypadków. Prawdopodobieństwo takiego zdarzenia jest równe $\frac{2}{6} = \frac{1}{3}$. Wyniki doświadczenia powinny zatem układać się według szeregu dwumianowego $\left(\frac{2}{3} + \frac{1}{3}\right)^{12}$. Widzimy, że zgodność teorii z faktami jest dosyć dobra. Rzuca się jednak w oczy, że odchylenia liczb teoretycznych od obserwowanych są z początku stale dodatnie, potem zaś stale ujemne. Jeżeli te odchylenia mają być rzeczą przypadku, to ujemne ich wartości powinny być rozrzucone wśród dodatnich. To nasuwa przypuszczenie, że prawdopodobieństwo rozpatrywanego zdarzenia nie było dokładnie ⅓. Przyczyną mogło być wykrzywienie kości, co przy tak wielkiej liczbie rzutów nie byłoby niespodzianką. Istotnie obrachowując liczbę wszystkich przypadków wystąpienia numeru 5 albo 6 otrzymujemy wartość 106 602 na ogólną liczbę przypadków wynoszącą 315 672. Daje to prawdopodobieństwo nieco większe od ⅓ a mianowicie 0.3376986. Obliczając ilości rzutów przy takim prawdopodobieństwie, otrzymujemy liczby kolumny piątej. Odchylają się one od liczb obserwowanych w mniejszym stopniu niż ilości rzutów obliczone dla prawidłowych kości, zestawione w kolumnie trzeciej. Co więcej, ujemne odchylenia są teraz pomieszane z dodatnimi. Można przypuścić, że pochodzą one z przypadku.

Rozpatrzone doświadczenie Fishera ma niemałe znaczenie dla zagadnienia prawa wielkich liczb (ust. 21). Widzimy tu dosyć znaczne odchylenie obserwowanych częstości od prawdopodobieństwa pomimo ogromnej liczby przypadków — z górą 300 tysięcy! Spotkaliśmy się nadto z trudnością, z jaką jest związane otrzymanie licznych materiałów statystycznych, z trudnością polegającą na ścisłym zachowaniu tych samych warunków w ciągu całego doświadczenia.

W doświadczeniu Fishera rozpatrywaliśmy po raz pierwszy stopień zgodności między szeregiem rozdzielczym a funkcją użytą do wyrównania. Ściślej mówiąc są tu dwa odrębne zagadnienia: jedno dotyczy stopnia zgodności między teorią a doświadczeniem, drugie zaś kwestii, czy odchylenia liczb doświadczenia od liczb teoretycznych pochodzą z przypadku, czy też z nieodpowiedniego doboru funkcji.

Pierwsze zagadnienie najprościej może być rozwiązane przez obliczenie stosunku sumy absolutnych wartości odchyleń od sumy wartości funkcji względnie liczebności szeregu. Ten stosunek w rozpatrzonym przykładzie przy założeniu prawidłowych kości wynosi 3.6%, przy założeniu zaś wykrzywionych 1.4% a więc z górą dwa razy mniej. Zgodność jest tym lepsza, im podany powyżej stosunek jest mniejszy.

Drugie zagadnienie jest bardziej zawiłe. Pearson zaproponował jego rozwiązanie przez użycie innego sposobu określenia zgodności między szeregiem a funkcją, mianowicie przez użycie funkcji χ^2, której wartości oblicza się w sposób następujący. Od częstości każdej wartości wariantów (f_i) obejmuje się odnośną wartość funkcji (y_i), różnicę podnosi się do kwadratu i dzieli się przez wartość funkcji. Otrzymane w ten sposób liczby dodaje się i dzieli przez wartość funkcji. Jest to wartość funkcji χ^2 dla danego szeregu obliczana według równania:

$$\chi^2 = \sum_{1}^{l} \frac{(f_i - y_i)^2}{y_i}$$

w którym l jest liczbą wartości wariantów, nie liczbą wariantów!

Im większa jest zgodność między szeregiem a funkcją, tym mniejsze jest χ^2. Następnie, mając wartość tej ostatniej funkcji dla danej próby, oblicza się, jaka część wszystkich możliwych prób wziętych z rozpatrywanej populacji da wartości równe i większe. Im większa część wypadnie, tym jest prób mniej zgodnych z funkcją niż próba dana. Jeżeli takich prób mniej zgodnych będzie dostatecznie dużo, będzie można

uznać, że funkcja została dobrana udatnie, w przeciwnym razie trzeba będzie wybrać inną funkcję. Mamy tu rozumowanie podobne do tego, z jakim spotkaliśmy się przy rozpatrywaniu wiarogodności charakterystyk dla małych prób (ust. 31). Przeprowadzenie tego rozumowania jest bardzo zawiłe i dlatego nie będę go podawał, odsyłając czytelnika do podręczników Fishera, Romanowskiego i Andersona.

Zgodność między szeregami a funkcjami będziemy nadal określali prostym sposobem podanym na początku. Do tego jeszcze dodam, że pewną charakterystykę zgodności daje także maksymalna wartość odchyleń a zwłaszcza rozkład znaków + i — w ciągu odchyleń: pomieszanie znaków świadczy o większej zgodności szeregu z funkcją niż rozdział ich. I jeszcze jedno: ważną jest rzeczą, żeby ciągi częstości obserwowanych i obliczonych „chwytały się końcami", to znaczy, żeby duża zgodność była dla najmniejszych i największych częstości. Dla pośrednich częstości można natomiast dopuścić większe różnice.

Jako drugi przykład wyrównywania szeregów rozdzielczych za pomocą szeregu dwumianowego, rozpatrzymy dane Geisslera co do ilości chłopców w niemieckich rodzinach mających ośmioro potomstwa. Materiał obejmował 53680 rodzin (tab. 34,2).

Tabela 34,2

Ilości chłopców	Liczby rodzin z takimi ilościami chłopców	Liczby te obliczone teoretycznie	Różnice
0	215	165	+50
1	1485	1402	+83
2	5331	5203	+128
3	10649	11035	−386
4	14959	14628	+331
5	11929	12410	−481
6	6678	6580	+98
7	2092	1994	+98
8	342	264	+78
Ogółem	53680	53681	−867 +866

W tym zagadnieniu mamy ogółem 221019 chłopców na 429440 dzieci. Zatem prawdopodobieństwo urodzenia chłopca wypada 0.51467, nieco większe niż urodzenia dziewczyny. Według tej wartości prawdopodobieństwa zostały obliczone liczby trzeciej kolumny za pomocą wzoru

$$5368\,(0.48533 + 0.51467)^8$$

Zostało to uskutecznione w przypuszczeniu, że płeć jednego dziecka jest niezależna od płci drugiego. Odchylenia liczb obserwowanych od obliczonych są na ogół tego samego rzędu wielkości co w poprzednim przykładzie. Ujemne wartości są pomieszane z dodatnimi. Wskazywałoby to na przypadkowość odchyleń, gdyby nie dziwne ujemne odchylenia o 3 i 5 chłopcach. Są to największe odchylenia i jedyne ujemne! Kryje się tu działanie jakiejś określonej przyczyny, zwłaszcza, że te ujemne odchylenia dotyczą wartości zmiennej ewentualnej nie byle jakich, lecz określonych — mianowicie położonych symetrycznie względem wartości modalnej. Pochodzi to być może z tej przyczyny, że powody w tej samej rodzinie nie są odnośnie do płci zdarzeniami niezależnymi od siebie. Stopień zgodności wynosi 3.2%.

35. Wyrównywanie szeregów rozdzielczych za pomocą normalnej funkcji prawdopodobieństwa. Spotyka się czasem szeregi rozdzielcze, dla których wskaźnik skośności drugiego rodzaju (γ_1) oraz eksces (γ_2) są małe. Takie szeregi można wyrównywać za pomocą normalnej funkcji prawdopodobieństwa (ust. 25).

Klasycznym przykładem są tu materiały statystyczne dotyczące wzrostu ludzi, np. materiał szwedzki przytoczony w ust. 6 (tab. 6,1). Wartości wzrostu są w nim zgrupowane w przedziałach centymetrowych. Trzeba je traktować tak, jak gdyby pomiary były wykonane w milimetrach. Do wyrównywania szereg jest za długi. Skrócimy go, łącząc klasy po dwie (tab. 35,1). Na tej podstawie otrzymamy charakterystyki: $m' = 86.115$, $\sigma = 2.9692$, $\gamma_1 = +0.1185 \pm 0.0113$, $\gamma_2 = -0.0425 \pm \pm 0.0226$.

Tabela 35,1

Wzrost Szwedów

Zakresy klas w *mm* x	Numery klas x'	Częstości f'
1510—	75.975	15
1530—	76.975	19
1550—	77.975	112
1570—	78.975	289
1590—	79.975	742
1610—	80.975	1458
1630—	81.975	2471
1650—	82.975	3818
1670—	83.975	5017
1690—	84.975	5896
1710—	85.975	6199
1730—	86 975	5881
1750—	87.975	5064
1770—	88.975	3887
1790—	89.975	2685
1810—	90.975	1645
1830—	91.975	920
1850—	92.975	476
1870—	93.975	223
1890—	94.975	89
1910—	95.975	45
1930—	96.975	22
1950—	97.975	7
1970—	98.975	1
	Ogółem	46981

Metoda porównywania szeregu rozdzielczego z normalną funkcją prawdopodobieństwa była już objaśniona w ust. 25 dla szeregu dwumianowego. Widzieliśmy, że są dwie jej odmiany: jedna za pomocą całki L a p l a c e'a przez porównywanie pól, druga za pomocą funkcji d e M o i v r e'a przez porównywanie rzędnych. W jednej i drugiej trzeba wprowadzić

do szeregu normalną formę dla zmiennej ewentualnej według równania

$$\xi = \frac{X' - m'}{\sigma'}$$

Przeprowadźmy z początku wyrównanie szeregu za pomocą całki Laplace'a. Trzeba będzie w tym celu wyznaczyć granice klas i po nadaniu formy normalnej dla zmiennej ewentualnej wybrać z tabeli *B* wartości całki dla wartości granic. Kolejne odejmowanie wartości dla całki da nam częstości względne dla poszczególnych klas. Mnożąc je przez liczebność szeregu otrzymamy teoretyczne częstości bezwzględne. Tok obliczeń jest zestawiony w tabeli 35,2, wyniki w tabeli 35,3. Z tej ostatniej wynika zgodność częstości obserwowanych z teoretycznymi w wymiarze 3,2%. Granice klas w tabeli 35.2 są oznaczone literą x.

Tabela 35,2

Porównanie wzrostu Szwedów z całką Laplace'a (Q)

x'	X	$X - m'$	$\xi = \frac{X-m'}{\sigma}$	Q	Częstości względne obliczone
75.975	75.475	—10.640	—3.58	0.0001	
					0.0005
76.975	76.475	—9.640	—3.25	0.0006	
					0.0012
77.975	77.475	—8.640	—2.91	0.0018	
					0.0033
	78.475	—7.640	—2.57	0.0051	
78.975					0.0075
	79.475	—6.640	—2.24	0.0126	
79.975					0.0161
	80.475	—5.640	—1.90	0.0287	
80.975					0.0307
	81.475	—4.640	—1.56	0.0594	

Tabela 35,2 (dokończenie)

Porównanie wzrostu Szwedów z całką Laplace'a (Q)

x'	X	$X-m'$	$\xi=\frac{X-m'}{\sigma}$	Q	Częstości względne obliczone
	81.475	—4.640	—1.56	0.0504	
81.975					0.0499
	82.475	—3.640	—1.23	0.1093	
82.975					0.0774
	83.475	—2.640	—0.89	0.1867	
83.975					0.1045
	84.475	—1.640	—0.55	0.2912	
84.975					0.1217
	85.475	—0.640	—0.22	0.4129	
85.975					0.1349
	86.475	+0.360	+0.12	0.5478	
86.975					0.1294
	87.475	+1.360	+0.46	0.6772	
87.975					0.1080
	88.475	+2.360	+0.79	0.7852	
88.975					0.0856
	89.475	+3.360	+1.13	0.8708	
89.975					0.0584
	90.475	+4.360	+1.47	0.9292	
90.975					0.0357
	91.475	+5.360	+1.81	0.9649	
91.975					0.0189
	92.475	+6.360	+2.14	0.9838	
92.975					0.0096
	93.475	+7.360	+2.48	0.9934	
93.975					0.0042
	94.475	+8.360	+2.82	0.9976	
94.975					0.0015
	95.475	+9.360	+3.15	0.9991	
95.975					0.0007
	96.475	+10.360	+3.49	0.9998	
96.975					0.0001
	97.475	+11.360	+3.83	0.9999	
97.975					0.0001
	98.475	+12.260	+4.16	1.0000	
98.975					0.0000
	99.475	+13.360	+4.50	1.0000	

Tabela 35,3

Porównanie wzrostu Szwedów z całką Laplace'a

Zakresy klas w *mm*	Częstości obserwowane	Częstości obliczone	Różnice
1510—	15	23	— 8
1530—	19	56	— 37
1550—	112	155	— 43
1570—	289	353	— 63
1590—	742	756	— 14
1610—	1458	1452	+ 16
1630—	5471	2345	+126
1650—	3818	3697	+181
1670—	5017	4916	+107
1690—	5896	5718	+178
1710—	6199	6338	—139
1730—	5881	6080	—199
1750—	5064	5074	— 10
1770—	3887	4022	—135
1790—	2685	2744	— 59
1810—	1645	1677	— 32
1830—	920	888	+ 32
1850—	476	451	+ 25
1870—	224	197	+ 26
1890—	89	70	+ 19
1910—	45	33	+ 12
1930—	22	5	+ 17
1950—	7	5	+ 2
1990—	1	0	+ 1
Ogółem	46981	46978	—739 +742

Przeprowadźmy następnie porównanie szeregu z normalną funkcją prawdopodobieństwa za pomocą funkcji de Moivre'a. Tak jak to jest pokazane w tabeli 35,4, nadajemy zmiennej ewentualnej formę normalną. Z tabeli *A* wybieramy odnośne wartości funkcji de Moivre'a. Będą to wartości zmiennej normalnej η, która jest związana z częstościami y równaniem $\eta = y\sigma$ Stąd wypada $y = \frac{\eta}{\sigma}$, a częstości bezwzględne obliczymy z równania $f = n\frac{\eta}{\sigma}$, gdzie n jest liczbą wariantów. Zgodność, jak to widać z tabeli 35,5 wypada równa 2.4%,

a więc lepsza niż przy stosowaniu całki Laplace'a. Zwykle bywa odwrotnie.

Tabela 35,4

Porównanie wzrostu Szwedów z funkcją de Moivre'a

x'	$x' - m'$	$\xi = \frac{x' - m'}{\sigma'}$	φ_0
75.975	—10.140	—3.42	0.0011
76.975	— 9.140	—3.08	0.0035
77.975	— 8.140	—2.74	0.0093
78.975	— 7.140	—2.40	0.0224
79.975	— 6.140	—2.07	0.0468
80.975	— 5.140	—1.73	0.0893
81.975	— 4.140	—1.39	0.1518
82.975	— 3.140	—1.06	0.2375
83.975	— 2.140	—0.72	0.3079
84.975	— 1.140	—0.38	0.3712
85.975	— 0.140	—0.05	0.3984
86.975	+ 0.860	+0.29	0.3825
87.975	+ 1.860	+0.63	0.3271
88.975	+ 2.860	+0.96	0.2516
89.975	+ 3.860	+1.30	0.1714
90.975	+ 4.860	+1.64	0.1040
91.975	+ 5.860	+1.97	0.0573
92.975	+ 6.860	+2.31	0.0277
93.975	+ 7.860	+2.65	0.0119
94.975	+ 8.860	+2.98	0.0047
95.975	+ 9.860	+3.32	0.0016
96.975	+10.860	+3.66	0.0005
97.975	+11.860	+3.99	0.0001
98.975	+12.860	+4.33	0.0000

Dla kontroli obliczeń dobrze jest dodawać częstości względne otrzymane z całki Laplace'a — suma powinna wypaść bliska jedności. W naszym przykładzie otrzymujemy 0.9999. Dla kontroli drugiej odmiany metody trzeba dodać wartości φ_0 funkcji de Moivre'a. Suma powinna być bliska wartości średniego odchylenia szeregu. W naszym przykładzie ta suma wynosi 2.9796 podczas gdy $\sigma' = 2.9692$.

Tabela 35,5

Porównanie wzrostu Szwedów z funkcją de Moivre'a

Zakresy klas w *mm*	Częstości obserwowane	Częstości obliczone	Różnice
1510—	15	17	— 2
1530—	19	55	— 36
1550—	112	147	— 35
1570—	289	353	— 64
1590—	742	740	+ 2
1610—	1458	1408	+ 50
1630—	2471	2394	+ 77
1650—	3818	3745	+ 73
1670—	5017	4855	+162
1690—	5896	5853	+ 43
1710—	6199	6282	— 83
1730—	5881	6031	—150
1750—	5064	5158	— 94
1770—	3887	3967	— 80
1790—	2685	2702	— 17
1810—	1645	1640	+ 5
1830—	920	903	+ 17
1850—	476	437	+ 39
1870—	223	188	+ 35
1890—	89	74	+ 15
1910—	45	25	+ 20
1930—	22	8	+ 14
1950—	7	2	+ 5
1970—	1	0	+ 1
Ogółem	46981	46984	—561 +558

36. Funkcje typu A. Jeżeli charakterystyki γ_1 i γ_2 szeregu rozdzielczego mają znaczne wartości, normalna funkcja prawdopodobieństwa nie nadaje się do wyrównywania. Jeżeli szereg jest dwuboczny, ale niezbyt silnie odbiega od formy normalnej, najlepiej jest użyć funkcyj Charlier'a typu *A*. Dla szeregów dwubocznych silnie odbiegających od formy normalnej trzeba natomiast użyć funkcyj typu *B*, tak jak dla szeregów jednobocznych. Będzie o tym mowa w następnym ustępie.

Funkcje Charliera typu A są to nieskończone sumy funkcyj, które przy użyciu współrzędnych normalnych mają formę:

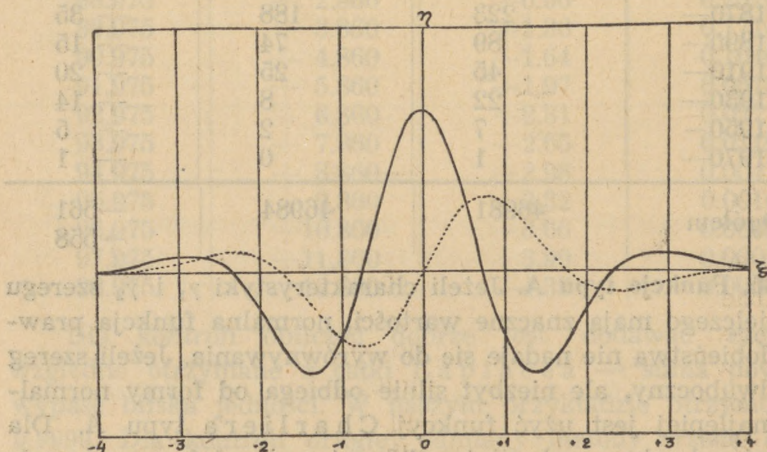
$$\varphi(\xi) = \varphi_0(\xi) + c_3\varphi_3(\xi) + c_4\varphi_4(\xi)$$

gdzie $\varphi_0(\xi)$ jest funkcją de Moivre'a a $\varphi_3(\xi)\ \varphi_4(\xi)\ \ldots$ są to pochodne tej ostatniej trzeciego, czwartego itd. rzędów. Wystarcza z powyższej sumy wziąć pierwsze trzy wyrazy. Rachunek nieskończonościowy pozwala określić formę pochodnych, które tu wchodzą, mianowicie:

$$\varphi_3(\xi) = \frac{d^3\varphi_0(\xi)}{d\xi^3} = (-\xi^3 + 3\xi)\,\varphi_0(\xi)$$

$$\varphi_4(\xi) = \frac{d^4\varphi_0(\xi)}{d\xi^4} = (\xi^4 - 6\xi^2 + 3)\,\varphi_0(\xi)$$

W końcu książki są podane w tabelach C i D wartości tych funkcyj dla wartości ξ w odstępach 0.01 z czterema znakami dziesiętnymi. Na razie przytaczam dla orientacji zestawione w tabeli 36,1 wartości $\varphi_3(\xi)$ i $\varphi_4(\xi)$ dla wartości ξ w odstępach 0.5. Rycina 9 przedstawia formę omawianych funkcyj.



Ryc. 9. Funkcje φ_3 (linia ciągła) i φ_4 (linia przerywana). Na osi odciętych zaznaczone wartości zmiennej niezależnej ξ. (Według Charliera).

Tabela 36,1

ξ	$\varphi_3(\xi)$	$\varphi_4(\xi)$
$-\infty$	0.0000	0.0000
—4.0	+0.0070	+0.0218
—3.5	+0.0288	+0.0694
—3.0	+0.0708	+0.1330
—2.5	+0.1424	+0.0800
—2.0	+0.1080	—0.2700
—1.5	—0.1457	—0.7043
—1.0	—0.4839	—0.4839
—0.5	—0.4841	+0.5501
0.0	0.0000	+1.1968
+0.5	+0.4841	+0.5501
+1.0	+0.4841	—0.4839
+1.5	+0.1457	—0.7043
+2.0	—0.1080	—0.2700
+2.5	—0.1424	+0.0800
+3.0	—0.0798	+0.1330
+3.5	—0.0283	+0.0694
+4.0	—0.0070	+0.0218
$+\infty$	0.0000	0.0000

Rachunek całkowy pozwala obliczyć dla funkcyj typu A średnią arytmetyczną i momenty drugiego, trzeciego i czwartego stopnia. Otrzymuje się:

$$m = 0$$
$$m_2 = 1$$
$$m_3 = -6C_3$$
$$m_4 = 24\,C_4 + 3$$

Dla porównania szeregu rozdzielczego z funkcją typu A trzeba wprowadzić do niego współrzędne normalne według równań $\xi = \frac{x-m}{\sigma}$, $\eta = \sigma\frac{f}{n}$ względnie $\xi = \frac{x'-m'}{\sigma'}$, $\eta = \sigma'\frac{f'}{n'}$.

Wtedy jego średnia arytmetyczna stanie się równa zeru, moment drugiego stopnia stanie się równy jedności, moment trzeciego stopnia przybierze wartość wskaźnika skośności drugiego rodzaju (γ_1) a moment czwartego stopnia — wartości ekscesu (γ_2) powiększonego o 3. W ten sposób ażeby charakterystyki funkcji typu A były równe odnośnym charakterystykom szeregu, współczynniki C_3 i C_4 powinny czynić zadość równaniom:

$$-6C_3 = \gamma_1$$

$$24\ C_4 + 3 = \gamma_2 + 3$$

Stąd dla nich otrzymujemy wzory:

$$C_3 = -\frac{1}{6}\gamma_1, \quad C_4 = +\frac{1}{24}\gamma_2.$$

Wprowadzenie współrzędnych normalnych jednocześnie przyrównuje średnią szeregu i jego moment drugiego stopnia odnośnym charakterystykom funkcji.

Porównywanie szeregu z funkcją może być przeprowadzone obu metodami użytymi w poprzednim ustępie. Wystarcza jednak porównywanie rzędnych. Unika się przez to złożonych rachunków, które nie przynoszą należytych korzyści.

Zastosujmy wyłożoną powyżej metodę do konkretnych materiałów statystycznych. Zaczniemy od wielokrotnie traktowanego materiału bławatka, zgrupowanego trzecim sposobem. Weźmiemy go w formie poprawionej na zasadzie kryterium Chauveneta bez krańcowego plus wariantu. Będziemy mieli dla niego charakterystyki: $m' = 11.547$, $\sigma' = 1.9053$, $\gamma_1 = +0.5088$, $\gamma_2 = +0.4464$ i otrzymamy $c_3 = -0.0848$, $c_4 = +0.0186$. Obliczmy wartości funkcji $\varphi_0(\xi)$ i $\varphi(\xi)$ dla poszczególnych wyrazów naszego szeregu (tab. 36,2). Widzimy z porównania częstości obliczonych i obserwowanych (tab. 36,3), że użycie funkcji $\varphi(\xi)$ dało wynik lepszy niż przy funkcji $\varphi_0(\xi)$. Wypada mianowicie stopień zgodności w pierwszym przypadku jest 11.2%, w drugim zaś 14.4%. O lepszej zgodności

Tabela 36,2

Liczby kwiatów w koszykach bławatka

x'	$x'-m'$	ξ	φ_0	φ_3	φ_4	$c_3\varphi_3$	$c_4\varphi_4$	φ
$7^1/_3$	—4.214	—2.21	0.0347	+0.1445	—0.0840	—0.0127	—0.0016	0.0208
$8^1/_3$	—3.214	—1.69	0.0957	+0.0233	—0.5720	+0.0020	—0.0107	0.0807
$9^1/_3$	—2.214	—1.16	0.2036	—0.3907	—0.6642	+0.0331	—0.0124	0.2243
$10^1/_3$	—1.214	—0.64	0.3251	—0.5389	+0.2309	+0.0457	+0.0043	0.3751
$11^1/_3$	—0.214	—0,11	0.3965	—0.1303	+1.1609	+0.0110	+0.0216	0.4291
$12^1/_3$	+0.786	+0.41	0.3668	+0.4259	+0,7408	—0.0361	+0.0138	0.3445
$13^1/_3$	+1.786	+0.94	0.2565	+0.5102	—0.3901	—0.0433	—0.0073	0.2059
$14^1/_3$	+2 786	+1.46	0.1374	+0.1742	—0.7209	—0.0148	—0.0134	0.1092
$15^1/_3$	+3.786	+1.99	0.0551	—0.1052	—0.2797	+0.0089	—0.0052	0.0588
$16^1/_3$	+4.786	+.251	0.0171	—0.1416	+0.0836	−0.0120	+0.0016	0.0307
$17^1/_3$	+5.786	+3.04	0.0040	—0.0746	+0.1290	+0.0063	+0.0024	0.0127
$18^1/_3$	+6.786	+3.56	0.0007	—0.0244	+0.0680	+0.0021	+0.0013	0.0041
			1.8932					1.8959

Tabela 36,3

Liczby kwiatów w koszykach bławatka

Zakresy klas	$\varphi_0 \frac{n}{\sigma'}$	$\varphi \frac{n}{\sigma'}$	f'	$\varphi_0 \frac{n}{\sigma^1} - f'$		$\varphi \frac{n}{\sigma'} - f'$	
21—	5.4	3.3	5	+ 0.4			— 1.7
24—	15.0	12.7	11	+ 4.0		+ 1.7	
27—	32.0	35.2	38		— 6.0		— 2.8
30—	51.0	58.9	61		—10.0		— 2.1
33—	62.2	67.3	63		— 0.8	+ 4.3	
36—	57.5	54.1	49	+ 8.5		+ 5.1	
39—	40.3	32.3	37	+ 3.3			— 4.7
42—	20.6	17.1	22		— 1.4		— 4.9
45—	8.6	9.2	5	+ 3.6		+ 4.2	
48—	2.7	4.8	4		— 1.3	+ 0.8	
51—	0.6	2.0	3		— 2.4		— 1.0
54—	0.1	0.6	1		— 0.9		— 0.4
Ogółem	296.0	297.5	299	+19.8	—22.8	+16.1	—17.6

przy użyciu funkcji $\varphi(\xi)$ świadczy także wymiar poszczególnych odchyleń, który wynosi najwyżej 5.1, podczas gdy przy funkcji $\varphi_0(\xi)$ dochodzi do 10.0. Zgodność jest zresztą mierna, co jest skutkiem małej liczebności próby. Według Charliera wogóle nie warto przeprowadzać wyrównywania szeregów przy mniej niż tysiącu wariantów w próbie, a tu było zaledwie 299.

Na przykładzie bławatka spróbujmy za pomocą wyrównania szeregu określić wartość modalną. Mieści się ona gdzieś w środku szeregu. Trzeba wobec tego obliczyć wartości funkcji $\varphi(\xi)$ dla poszczególnych najczęstszych wartości zmiennej ewentualnej. W tabeli 36,4 jest przedstawiony tok obliczeń. Wypada z nich, że najczęstszą jest wartość $x = 34$, a więc liczba z szeregu Fibonacciego. Nie jest to wynik pewny z uwagi na małą liczebność materiału.

Tabela 36,4

Liczby kwiatów w koszykach bławatka

x	x'	$x'-m'$	ξ	φ_0	φ_3	φ_4	$c_3\ \varphi_3$	$c_4\ \varphi_4$	φ
31	$10^1/_3$	—1.214	—0.64	0.3251	—0.5389	+0.2309	+0.0457	+0.0043	0.3751
32	$10^2/_3$	—0.880	—0.46	0.3589	—0.4603	+0.6371	+0.0390	+0.0119	0.4098
33	11	—0.547	—0.29	0.3825	—0.3235	+0.9572	+0.0274	+0.0178	0.4277
34	$11^1/_3$	—0.214	—0.11	0.3965	—0.1303	+1.1609	+0.0110	+0.0216	0.4291
35	$11^2/_3$	+0.120	+0.06	0.3982	+0.0716	+1.1861	—0.0061	+0.0221	0.4142
36	12	+0.453	+0.24	0.3876	+0.2737	+1.0302	—0.0232	+0.0192	0.3836
37	$12^1/_3$	+0.786	+0.41	0.3668	+0.4259	+0.7408	—0.0361	+0.0138	0.3445

Daleko lepsze wyniki wyrównania da drugi materiał, jaki rozpatrzymy — materiał dotyczący wskaźnika głównego szwedzkich rekrutów z lat 1897 i 1898, ogłoszony w r. 1902 przez Petriusa i Fürsta (tab. 36,5). Zapożyczam go z książki Charliera. Mamy tu 22 505 wariantów.

Wskaźnik główny jest to stosunek szerokości czaszki do jej długości. Wyraża się go w setnych częściach.

Tabela 36,5

Wskaźnik główny szwedzkich rekrutów (setne części jedności)

Zakresy klas	Ich częstości
65—66	12
67—68	87
69—70	510
71—72	1952
73—74	4346
75—76	6039
77—78	5050
79—80	2822
81—82	1172
83—84	377
85—86	94
87—88	31
89—90	13
Ogółem	22505

Charakterystyki tego drugiego szeregu są następujące: $m' = 38.029$, $\sigma' = 1.5176$, $\gamma_1 = +0.2553$ $\gamma_2 = +0.3901$. Stąd otrzymamy: $c_3 = -0.0426$, $c_4 = +0.0163$. W tabelach 36,6—7 są podane szczegóły rachunków. Tabela 36,8 daje zestawienie częstości obserwowanych i obliczonych. Wypływa z niej że stopień zgodności wynosi przy użyciu funkcji $\varphi(\xi)$ 3.4%, przy użyciu zaś funkcji $\varphi_0(\xi)$ 5.5%.

Tabela 36,6

Wskaźnik główny szwedzkich rekrutów

Zakresy klas	x'	$x' - m'$	$\xi = \dfrac{x' - m'}{\sigma'}$	φ_0
65—66	32.75	—5.279	—3.48	0.0010
67—68	33.75	—4.279	—2.82	0.0075
69—70	34.75	—3.279	—2.16	0.0387
71—72	35.75	—2,279	—1.50	0.1295
73—74	36.75	—1.279	—0.84	0.2803
75—76	37.75	—0.279	—0.18	0.3925
77—78	38.75	+0.721	+0.48	0.3555
79—80	39.75	+1.721	+1.13	0,2107
81—82	40.75	+2.721	+1.79	0.0804
83—84	41.75	+3.721	+2.45	0,0198
85—86	42.75	+4.721	+3.11	0.0032
87—88	43.75	+5.721	+3.77	0.0003
89—90	44.75	+6.721	+4.43	0.0000

Tabela 36,7

Wskaźnik główny szwedzkich rekrutów

ξ	φ_3	$c_3\varphi_3$	φ_4	$c_4\varphi_4$	φ
—3.48	+0.0298	—0.0013	+0.0721	+0.0012	0.0009
—2.82	+0.1045	—0.0044	+0.1386	+0.0023	0.0054
—2.16	+0.1393	—0.0059	—0.1249	—0.0020	0.0308
—1.50	—0.1457	+0.0062	—0.7043	—0.0114	0.1243
—0.84	—0.5403	+0.0230	—0.2063	—0.0034	0.2999
—0.18	—0.2097	+0.0089	+1.1017	+0.0179	0.4193
+0.48	+0.4726	—0.0201	+0.5940	+0.0097	0.3451
+1.13	+0.4102	—0.0175	—0.6386	—0.0104	0.1828
+1.79	—0.0294	+0.0012	—0.4789	—0.0078	0.0738
+2.45	—0.1459	+0.0062	—0.0598	+0.0010	0.0270
+3.11	—0.0657	+0.0028	+0.1219	+0.0010	0.0080
+3.77	—0,0139	+0.0006	+0 0394	+0.0006	0.0018
+4.43	—0.0016	+0.0001	+0.0060	+0.0001	0.0002

Tabela 36,8

Wskaźnik główny szwedzkich rekrutów

Zakresy klas	Częstości obserwowane f'	Częstości obliczone $\varphi_0 \frac{n}{\sigma'}$	Częstości obliczone $\varphi \frac{n}{\sigma'}$	Różnice $f' - \varphi_0 \frac{n}{\sigma'}$	Różnice $f' - \varphi \frac{n}{\sigma'}$
65—	12	15	13	− 3	− 1
67—	87	111	80	− 24	+ 7
69—	519	573	456	− 64	+ 54
71—	1952	1918	1841	+ 34	+111
73—	4346	4152	4442	+194	− 90
75—	6039	5814	6211	+225	−172
77—	5050	5266	5119	−216	− 69
79—	2822	3121	2708	−299	+114
81—	1172	1191	1093	− 19	+ 79
83—	377	293	400	+ 84	− 23
85—	94	47	119	+ 47	− 25
87—	31	4	27	+ 27	+ 4
89—	13	0	3	+ 13	+ 10
Ogółem	22505	22505	22512	—624 +624	−386 +379

Jako trzeci przykład przytaczam (tab. 36,9) szereg opracowany przez J. Witkowskiego, dotyczący błędów obserwacyjnych rektascencji gwiazd. (tab. 6,8). Charakterystyki są tu następujące: $m' = +0.133$, $\sigma' = 1.967$, $c_3 = +0.004$, $c_4 = -0.017$. Były one obliczone bez poprawki Shepparda. Zgodność częstości obserwowanych i obliczonych wynosi 4.7%.

Dodatkowo warto jest zauważyć, że opracowany w poprzednim ustępie materiał wzrostu Szwedów wymaga właściwie zastosowania funkcji typu *A*. Wskazuje na to rozkład znaków przy różnicach między częstościami teoretycznymi i obserwowanymi. Pozostawiam czytelnikowi sprawdzenie.

Tabela 36,9

Błędy obserwacyj rektascencji gwiazd

Wartość środkowa klas	Częstości		Różnice
	obserwowane	obliczone	
—0.90	9	6	+ 3
—0.75	55	72	— 17
—0.60	245	289	— 44
—0.45	854	741	+113
—0.30	1358	1386	— 39
—0.15	1948	1983	— 35
0.00	2297	2374	— 77
+0.15	2052	2121	— 69
+0.30	1582	1585	— 3
+0.45	1020	910	+110
+0.60	351	380	— 29
+0.75	70	102	— 32
+0.90	13	12	+ 1
Ogółem	11854	11961	—333 +227

37. Funkcje typu B. Do wyrównywania szeregów jednobocznych a także dwubocznych silnie odbiegających od formy normalnej trzeba uciec się do funkcji Charliera typu *B*. Wywodzą się one z szeregu Poissona, stanowiąc pewnego rodzaju uogólnienie tego szeregu. Ogólny wzór tych funkcyj ma formę nieskończonej sumy:

$$\psi(x) = \psi_0(x) + k_2 \Delta^2 \psi_0(x) + k_3 \Delta^3 \psi_0(x) + \dots$$

W nim $\psi_0(x)$ jest to znana już nam z ust. 26 funkcja Poissona

$$\psi_0(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$k_2, k_3, \dots$ są to stałe spółczynniki a $\Delta^2 \psi_0(x)$, $\Delta^3 \psi_0(x) \dots$ są przyrostami funkcji drugiego, trzeciego itd. stopnia.

Oblicza się te przyrosty według wzorów:

$$\Delta \psi_0 (x) = \psi_0(x) - \psi_0(x-1)$$

$$\Delta^2 \psi_0' x = \Delta \psi_0 x - \Delta \psi_0 x-1 = \psi_0 x - 2 \psi_0' x-1) + \psi_0' x-2)$$

$$\Delta^3 \psi_0(x) = \Delta^2 \psi_0(x) - \Delta^2 \psi_0(x-1) = \psi_0(x) - 3 \psi_0(x-1) +$$
$$+ 3 \psi_0(x-2) - \psi_0(x-3)$$

. . . .

Z sumy określającej funkcję typu *B* bierze się ograniczoną liczbę wyrazów. Zależnie od formy szeregu rozdzielczego, który ma być wyrównany, bierze się jeden, dwa lub trzy wyrazy. Charlier brał dwa wyrazy. Dodanie trzeciego wyrazu nie było dotychczas, o ile mi jest wiadome, przez nikogo stosowane. [1])

Według zasady wyłożonej w ust. 33 funkcja, która ma służyć do wyrównania szeregu, powinna mieć średnią i momenty równe odnośnym charakterystykom szeregu. W zastosowaniu do rozpatrywanego zagadnienia będzie tu chodziło poza średnią o momenty drugiego i trzeciego stopnia. Moment czwartego stopnia byłby potrzebny tylko wtedy, gdyby się wzięło czwarty wyraz w nieskończonej sumie $\psi(x)$. Ale to jest bezcelowe.

Dla funkcyj typu *B* otrzymuje się metodą, zastosowaną w ust. 26 do szeregu Poissona, następujące charakterystyki:

$$m = \lambda$$
$$m_2 = \lambda + 2k_2$$
$$m_3 = \lambda + 6k_2 - 6k_3$$

Stąd otrzymujemy dla charakterystycznych stałych rozpatrywanych funkcyj następujące wzory:

$$\lambda = m$$

$$k_2 = \frac{m_2 - m}{2}$$

$$k_3 = \frac{3m_2 - m_3 - 2m}{6}$$

---

[1]) Zob. pracę autora tej książki pt. „Observations biométriques. VI" (Acta Societatis Botanicorum Poloniae. Tom XVII (1946).

Przy obliczeniach wartości funkcji typu *B* trzeba pamiętać o tym, że dla ujemnych wartości zmiennej niezależnej funkcja Poissona ma stale wartość zero.

Przy stosowaniu funkcyj typu *B* do wyrównywania szeregów jednobocznych nie trzeba wprowadzać współrzędnych normalnych. Trzeba natomiast przenieść początek współrzędnych dla zmiennej ewentualnej tak, żeby wartość jej mająca największą częstość równała się zeru. I tak dla liczby płatków w szczytowych kwiatach jaskra płożącego się (*Ranunculus repens*) mamy szereg (tab. 6,10):

5	6	7	8	9	10
314	56	26	13	3	1

Nadajemy mu formę:

0	1	2	3	4	5
314	56	26	13	3	1

Funkcje typu *B* dają częstości względne, które przerachowujemy na bezwzględne mnożąc je przez liczbę wariantów.

Przeprowadźmy wyrównanie przytoczonego powyżej szeregu jaskra płożącego się. Obliczamy jego charakterystyki: $m = 0.3971$, $m_2 = 0.6898$, $m_3 = +1.3721$. Stąd wypada $\lambda = 0.3971$, $e^{-\lambda} = 0.6723$, $k_2 = +0.1463$, $k_3 = -0.0161$. W tabeli 37,1 są zestawione przeliczenia w dziesięciotysięcznych. Przerachowując otrzymane częstości względne na bezwzględne ($n = 413$), otrzymujemy zestawienie wartości teoretycznych i obserwowanych (tab. 37,2). Zgodność wypadła doskonała, stopień jej jest 0.08%.

W wielu przypadkach trzeba pominąć trzeci wyraz funkcji $\psi(x)$. Na przykład dla liczby płatków u przylaszczki (*Anemone Hepatica*) (tab. 6,10) lepszy wynik otrzymuje się z dwoma wyrazami niż z trzema. Pozostawiam czytelnikowi przeprowadzenie obliczeń.

Tabela 37,1

Liczby płatków w szczytowych kwiatach jaskra płożącego się

x	ψ_0	$\Delta\psi_0$	$\Delta^2\psi_0$	$\Delta^3\psi_0$	$k_2\Delta^2\psi_0$	$k_3\Delta^3\psi_0$	ψ
0	6723	+6723	+ 6723	+ 6723	+ 984	−108	7599
1	2670	−4053	−10776	−17499	−1577	+282	1375
2	530	−2140	+ 1913	+12689	+ 280	−204	606
3	70	− 460	+ 1680	− 233	+ 246	+ 4	320
4	7	− 63	+ 397	− 1283	+ 58	+ 21	86
5	0	− 7	+ 56	− 341	+ 8	+ 5	13
	10000				−1577 +1576	−312 +312	9999

Tabela 37,2

Liczby płatków u jaskra płożącego się

Liczba płatków	Ich częstości		Różnice
	obliczone	obserwowane	
5	313.8	314	—0.2
6	56.8	56	+0.8
7	25.0	26	—1.0
8	13.2	13	—0.2
9	3.6	3	+0.6
10	0.5	1	—0.5
Ogółem	412.9	413	—0.17 +0.16

Czasem dla wyrównania szeregu trzeba ograniczyć się do pierwszego wyrazu funkcji $\psi(x)$, to znaczy wziąć po prostu funkcję Poissona. Klasycznym tego przykładem jest materiał wypadków śmierci od kopnięcia konia w armii pruskiej, opracowany przez Bortkiewicza (tab. 6,13). Tu $m = \lambda = 0.61$, $e^{-\lambda} = 0.5433$. W tabeli 37,3 są zestawione częstości.

Wymiar zgodności wynosi 2.4%. Jest ona gorsza niż w poprzednich dwóch wypadkach.

Tabela 37,3

Wypadki śmierci od kopnięcia konia w 20 korpusach armii pruskiej w ciągu 10 lat

x	Częstości		Różnice
	obliczone	obserwowane	
0	108.7	109	—0.3
1	66.3	65	+1.3
2	20.2	22	—1 8
3	4.1	3	+1.1
4	0.6	1	—0.4
Ogółem	199.9	200	—2.5 +2.4

Jak to było wzmiankowane na początku tego ustępu, funkcje Charliera typu B mogą być także użyte do wyrównywania szeregów dwubocznych, jeżeli te szeregi silnie odbiegają od formy normalnej. Mamy dwa takie przypadki: w jednym szeregi są silnie skośne, w drugim poza skośnością, która zresztą bywa różna, częstość wartości modalnej znacznie przewyższa częstości sąsiednich wartości zmiennej ewentualnej.

Dla ilustracji pierwszego rodzaju takich szeregów dwubocznych przytoczę przykład Charliera, jedyny zresztą, jaki przez niego został podany w jego podręczniku. Dotyczy on częstości dni z burzami w sierpniu w Lund (tab. 6,12). Dla niego $\lambda = 2.133$, $k_2 = 1004$ — użyte tu są dwa wyrazy funkcji $\psi(x)$. W tabeli 37,4 są zestawione częstości obserwowane i obliczone. Zgodność wypadła mierną: wymiar jej wynosi 13.6% a oprócz tego ciągi częstości nie bardzo się „trzymają końcami".

Tabela 37,4

Dni z burzami w sierpniu w Lund

Liczby dni	Częstości		Różnice
	obserwowane	obliczone	
0	24	24.8	−0.8
1	26	28.2	−2,2
2	19	15.9	+3.1
3	13	10.2	+2.8
4	9	9.5	−0.5
5	6	7.9	−1.9
6	5	4.6	+0.4
7	2	2.5	−0.5
8	0	0.7	−0.7
9	0	0.3	−0.3
10	0	0.1	−0.1
11	1	0.0	+1.0
Ogółem	105	104.7	−7.0 +7.3

Jako drugi przykład, przytoczę ciekawy materiał Studenta podany przez Fishera w jego podręczniku. Obserwowano pod mikroskopem zawiesinę drożdży i liczono komórki, które znalazły się w 400 małych kwadratach, na jakie był podzielony *1 mm*² na szkiełku przedmiotowym. Jest to zasada, na której opiera się użycie hematocymetru, przyrządu służącego do określenia liczby ciałek krwi. W tabeli 37,5 są zestawione obserwowane i obliczone częstości poszczególnych liczb komórek. Obliczenia przeprowadzono za pomocą samej tylko funkcji Poissona, dla której w tym przypadku $\lambda = 4.68$, $e^{-\lambda} = 0.009279$. Zgodność wypadła mierna. Jej wymiar, równy 9.6%, jest wprawdzie lepszy niż w poprzednim przypadku, ale „trzymanie się końcami” przedstawia się gorzej.

Tabela 37,5

Komórki drożdży obserwowane w hematocymetrze

Liczby komórek	Ich częstości		Różnice
	obserwowane	obliczone	
0	0	3.7	− 3.7
1	20	17.4	+ 2.6
2	43	40.6	+ 2.4
3	53	63.4	−10.4
4	86	74.2	+11.8
5	70	69.4	+ 0.6
6	54	54.2	− 0.2
7	37	36.2	+ 0.8
8	18	21.2	− 3.2
9	10	11.0	− 1.0
10	5	5.2	− 0.2
11	2	2.2	− 0.2
12	2	0.9	+ 1.1
Ogółem	400	399.6	−18.9 +19.3

Przechodzimy w końcu do rozpatrzenia wyrównywania szeregów dwubocznych odznaczających się wybitną częstością wartości modalnej, poza tym zresztą zwykle dosyć silnie asymetrycznych. W ust. 6, 7 i 19 były przytoczone przykłady tego rodzaju szeregów. Między innymi taki szereg daje meduza *Pseudoclytia pentata* (tab. 6,5). Zastosowanie do nich funkcyj typu *B* wymaga gruntownej przeróbki odnośnych szeregów rozdzielczych. Trzeba tu rozumować w sposób następujący. W odnośnych zjawiskach są czynne trzy rodzaje przyczyn. Pierwsza z nich działa stale i silniej od innych, powoduje przeto wyjątkową częstość wartości modalnej. Dwie pozostałe działają rzadziej i słabiej. Jedna z nich powoduje powiększenie zmiennej ewentualnej, druga zaś — pomniejszenie. Można wobec tego z danego szeregu wydzielić dwa nowe szeregi. Jeden z nich będzie charakteryzował działanie przyczyny pomniejszającej zmienną ewentualną a drugi — działanie przyczyny powiększającej tę zmienną.

Otóż dla liczb promieni u meduzy *Pseudoclytia* mieliśmy szereg rozdzielczy:

Liczby promieni:	2	3	4	5	6	7	8
Ich częstości:	1	8	56	860	64	6	1

Dla przyczyny pomniejszającej liczbę promieni wypada szereg:

Działania przyczyny:	0	1	2	3
Ich częstości:	931	56	8	1

Weźmiemy w tym przypadku dwa wyrazy funkcji $\psi(x)$. Otrzymamy: $m = 0.0753$, $m_2 = 0.0917$, $\lambda = 0.0753$, $e^{-\lambda} = 0.9275$, $k_2 = +0.0082$. Zgodność wypada w wymiarze 0.3% (tab. 37,6) a więc bardzo dobra.

Tabela 37,6

Przyczyna pomniejszająca liczbę promieni meduzy

Działania przyczyny	Ich częstości		Różnice
	obliczone	obserwowane	
0	931.4	931	+0.4
1	55.0	56	−1.0
2	9.1	8	+1.1
3	0.6	1	−0.4
Ogółem	996.1	996	−1.4 +1.5

Dla przyczyny powiększającej liczbę promieni otrzymamy szereg:

Działania przyczyn:	0	1	2	3
Ich częstości:	925	64	6	1

Zastosujemy tu trzy wyrazy funkcji $\psi(x)$. Charakterystyki wypadną: $m = 0.0793$, $m_2 = 0.0911$, $m_3 = 0.1170$, $k_2 = +0.0059$, $k_3 = -0.0004$. Zgodność wypada w wymiarze 0.2% (tab. 37,7) a więc doskonała.

Tabela 37,7

Przyczyna powiększająca liczbę promieni meduzy.

Działania przyczyny	Ich częstości		Różnice
	obliczone	obserwowane	
0	925.3	925	+0.3
1	63.7	64	−0.3
2	6.2	6	+0.2
3	0.1	1	−0 9
Ogółem	995.3	996	−1.2 +0.5

Przy zastosowaniu funkcyj *B* wypadają czasem rażące niezgodności między częstościami obliczonymi a obserwowanymi, powodowane przez skąpe warianty. Na przykład dla jaskra wielokwiatowego *(Ranunculus polyanthemos)* były przytoczone następujące częstości dla liczb płatków (tab. 6,10).

Liczby płatków:	5	6	7	8	9
Ich częstości:	410	16	1	0	1

Charakterystyki wypadną: $m = 0.05140$, $m_2 = 0.08148$, $m_3 = 0.19293$, $\lambda = 0.05140$, $e^{-\lambda} = 0.9499$, $k_2 = +0.01504$, $k_3 = -0.0855$. W konsekwencji otrzymuje się częstości paradoksalne (tab. 37,8) — jedna z nich ujemna!

Tabela 37,8

Liczby kwiatów płatków u *Ranunculus polyanthemos.*

Liczby płatków	Ich częstości	
	obliczone	obserwowane
5	409.2	410
6	19.2	16
7	−4.3	1
8	3.3	0
9	0.2	1
Ogółem	427.6	428

Jest to spowodowane przez napotkanie bardzo rzadkiego osobnika rośliny z 9 płatkami. Wystarczy odrzucić ten wyjątkowy wariant, by otrzymać doskonałą zgodność, jak to widać z tabeli 37,9. Charakterystyki teraz są: $m = 0.04215$, $m_2 = 0.04502$, $m_3 = 0.05044$, $\lambda = 0.04215$. $e^{-\lambda} = 0.9585$, $k_2 = +0.00143$, $k_3 = +0.00005$. Mamy tu nowy przyczynek do zagadnienia skrajnych wariantów, omówionego w ust. 32.

W końcu dodatkowo trzeba jeszcze powiedzieć parę słów o wyrównywaniu szeregów wklęsłych, występujących prawie wyłącznie przy badaniach nad zachmurzeniem (por. tab. 6,14). Żadna z omówionych w tym rozdziale funkcyj nie nadaje się dla tej formy szeregów rozdzielczych. Pearson, układając funkcje do wyrównywania szeregów, nie pominął tej wyjątkowej formy zmienności. Ponieważ to nie ma poważniejszego znaczenia, nie będę się tą kwestią zajmował. Czytelnik znajdzie odnośną funkcję Pearsona w książce Eldertona (zob. ust. 33).

Tabela 37,9

Liczby płatków u *Ranunculus polyanthemos*.

Liczby płatków	Ich częstości	
	obliczone	obserwowane
5	409.9	410
6	16.1	16
7	1.1	1
Ogółem	427.0	427

CZĘŚĆ DRUGA

WSPÓŁZALEŻNOŚĆ MIĘDZY POPULACJAMI

ROZDZIAŁ VII.

KORELACJA

38. Pojęcie korelacji. Współzależność między populacjami może mieć charakter korelacji i kontyngencji. Pierwsze zachodzi wtedy, kiedy warianty różnią się między sobą ilościowo, drugie — jeżeli różnice między wariantami w obu porównywanych populacjach albo przynajmniej w jednej z nich mają charakter jakościowy.

Można porównywać ze sobą także nie dwie ale liczne populacje. Wtedy zjawia się zagadnienie klasyfikacji, czego rozpatrywać nie będziemy.

Jako przykłady korelacji można przytoczyć współzależność między liczbą kwiatów języczkowych i rurkowych w koszykach roślin z rodziny złożonych *(Compositae)*, współzależność między wzrostem ojców i synów, między wiekiem męża i żony itd. Jako przykłady kontyngencji można podać współzależność między barwą oczu a barwą włosów, między wypadkami śmierci od dyfterytu a stosowaniem szczepionki Behringa, między śmiertelnością a płcią itp.

W tym rozdziale zajmiemy się zagadnieniem korelacji, odkładając kontyngencję do następnego rozdziału. W zagadnieniu korelacji mamy do czynienia z dwoma populacjami, zazwyczaj z próbkami dwóch populacyj o jednakowej liczbie wariantów, przy czym każdemu wariantowi jednej odpowiada pewien wariant drugiej, dla każdego z nich inny, jakkolwiek niekoniecznie

odmienny. Takie powiązanie wariantów dwóch populacyj pochodzi bezpośrednio albo pośrednio z bliskości w czasie lub w przestrzeni przedmiotów czy zjawisk będących źródłem rozpatrywanych populacyj. Naprzykład korelacja między liczbami kwiatów rurkowych i języczkowych pochodzi stąd, że jedne i drugie bierze się z tych samych koszyków, korelacja między temperaturami dwóch miast pochodzi z jednoczesności obserwacyj meteorologicznych. O ile chodzi o pośredni stosunek przestrzenny i czasowy, można przytoczyć korelację między wzrostem ojców i synów, między wiekiem męża i żony. Czy takie związki są przestrzenne czy pośrednie, zawsze istotnym ich źródłem są związki przyczynowe. Nie wynika z tego bynajmniej, by statystyka miała zajmować się badaniem przyczynowości. Celem jej zawsze tylko opis, który zresztą może być użyteczny do badań nad przyczynowością. Stąd pochodzi równorzędne traktowanie obu porównywanych populacyj w zagadnieniach korelacji. Na przykład statystyka rozpatruje równorzędnie zależności wzrostu synów od wzrostu ojców i zależność wzrostu ojców od wzrostu synów, jakkolwiek wzrost synów jest częściowo uwarunkowany przez wzrost ojców, natomiast na wzrost ojców wzrost ich synów nie ma żadnego wpływu. Przy rozpatrywaniu korelacji, tak samo zresztą jak kontyngencji, chodzi tylko o związki matematyczne, funkcjonalne a nie przyczynowe. Dlatego właśnie mówi się o współzależności między populacjami a nie o zależności jednej od drugiej.

Ustalmy teraz dokładniej zadanie badania korelacyjnego. Podałem powyżej, że każdemu wariantowi jednej populacji odpowiada pewien wariant drugiej, dla każdego inny, ale niekoniecznie odmienny. Z reguły przy tym jest tak, że danej wartości wariantów jednej populacji odpowiadają różne i to liczne wartości wariantów drugiej. To nas prowadzi do rozpatrywania związków nie pomiędzy poszczególnymi wariantami, lecz między wartościami wariantów. Trzeba znowu zrobić użytek z pojęcia zmiennej ewentualnej, ale będziemy tu mieli nie

jedną jak poprzednio zmienną tego rodzaju, lecz dwie — osobne w rozpatrywanych populacjach.

Niech będzie populacja, względnie próba z populacji, złożona z n wariantów $u_1, u_2, \dots u_n$. Przypuśćmy, że mają one h różnych wartości $x_1, x_2, \dots, x_h$, przy czym naturalnie $h < n$. Niech będzie druga populacja złożona tak jak pierwsza z n wariantów $v_1, v_2, \dots, v_n$ wykazujących l różnych wartości $y_1, y_2, \dots, y_l$, gdzie $l < n$ i nadto z reguły $l \neq h$. Niech każdemu wariantowi u_i pierwszej populacji odpowiada wariant v_i drugiej. Zmienna x będzie zmienną ewentualną pierwszej populacji, zmienna y takąż zmienną drugiej.

Chcąc teraz przedstawić za pomocą jakiejś funkcji związek między populacjami, nie będziemy tego mogli zrobić za pomocą samych tylko zmiennych x i y w formie $y = f_1(x)$, $x = f_2(y)$, bo każdej danej wartości zmiennej x będą naogół odpowiadały różne wartości zmiennej y i odwrotnie. Trzeba będzie wprowadzić nowe zmienne X i Y a to w sposób następujący. Zmienna Y ma dla każdej wartości x przybierać wartość pośrednią między odpowiadającymi jej wartościami zmiennej y. Będzie to pewna funkcja zmiennej $x : Y = f_1(x)$. W podobny sposób dla zmiennej x trzeba ustalić nową zmienną X związaną ze zmienną y równaniem $X = f_2(y)$. Warto jest przy tym zrobić uwagę, że funkcja f_2 nie jest odwrotnością funkcji f_1. Przytoczone powyżej równanie dla zmiennych Y i X noszą nazwę równań regresji. Termin regresji pochodzi stąd, że badanie korelacji zrodziło się z badań nad dziedzicznością. Mianowicie zapoczątkował je angielski biolog i etnograf Franciszek Galton w dziele pod tytułem „Natural Inheritance”, które ukazało się w r. 1889. Powodem wyboru tego terminu było stwierdzenie faktu, że potomstwo na ogół mniej odbiega od przeciętności niż jego rodzice, ulega więc uwstecznieniu — regresji, co ujawnia się przy badaniu korelacji między cechami potomstwa i rodziców Warto jest jeszcze zaznaczyć, że matematyczne podstawy ra-

chunku korelacyjnego wywodzą się z jednej z prac francuskiego fizyka A u g u s t a B r a v a i s (1811—1863), która ukazała się w r. 1846.

Jeszcze trzeba dodać jedno wyjaśnienie do pojęcia korelacji. W praktyce nigdy prawie nie ma się do czynienia z całymi populacjami, tylko z próbami, nieraz o małej liczebności. Otóż jeżeli próby są małe, często się zdarza, że poszczególnym wartościom jednej zmiennej ewentualnej odpowiadają pojedyncze wartości drugiej. Zagadnienie jednak pozostaje bez zmiany, gdyż zadaniem rachunku korelacyjnego jest badanie związku nie między próbami, lecz całymi populacjami. Przy małych próbach możnaby wprawdzie ustalić związek funkcjonalny między zmiennymi ewentualnymi za pomocą wzoru interpolacyjnego L a g r a n g e'a, ale to nie dałoby żadnego obrazu związku między populacjami. Trzeba i w tym wypadku wprowadzić owe zmienne Y i X i użyć równań regresji takiej samej formy, jak przy traktowaniu dużych prób. Nie można będzie wtedy powiedzieć, że zmienna Y przybiera dla każdej wartości x wartość pośrednią między odpowiadającymi jej wartościami zmiennej y w rozpatrywanym materiale statystycznym, bo w nim jest tylko jedna taka wartość. Zmienna Y będzie jednak przybierać wartości pośrednie między wartościami y, ale między wartościami całości populacji, nie danego materiału.

Forma funkcyj $f_1(x)$ i $f_2(y)$ może być różna i da się ogólnie przedstawić w formie parabolicznej:

$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3 + \ldots..$$

$$f_2(y) = a_2 + b_2 y + c_2 y^2 + d_2 y^3 + \ldots.$$

W praktyce statystycznej wystarcza przeważnie skrócona forma liniowa

$$f_1(x) = a_1 + b_1 x$$

$$f_2(y) = a_2 + b_2 y$$

i rachunki sprowadzają się do wyznaczenia dwu tylko współczynników dla każdej funkcji.

Taka prosta forma równań regresji ma swoje źródło w tym, że zmienność wariantów w populacjach jest skutkiem działania licznych przyczyn, z których każda z osobna nie wywołuje wybitnych skutków. Jest to ten sam stan rzeczy, który powoduje, że zmienność da się przedstawić za pomocą określonych funkcyj, wywodzących się z szeregu dwumianowego. Uwzględnienie bardziej złożonych form równań regresji wymaga bardzo zawiłych rachunków i przeto nie może być omawiane w tej książce.

39. Wyznaczanie stałych równania regresji. Zadanie to może być rozwiązane w rozmaity sposób. Jednakże jedno wymaganie musi być zawsze uwzględnione: mianowicie suma dodatnich odchyleń wariantów y od odnośnych wartości Y ma być równa sumie odchyleń ujemnych, i tak samo suma dodatnich odchyleń x od X ma być równa sumie odchyleń ujemnych.

Najbardziej naturalnym rozwiązaniem rozpatrywanego zagadnienia byłoby znalezienie takich wartości dla współczynników a i b, żeby wspomniane powyżej sumy były możliwie małe. Niestety, dla takiego rozwiązania nie można wyprowadzić żadnych ogólnych wzorów, a w poszczególnych przypadkach trzeba by było wyszukiwać wartości współczynników przez mozolne próby, po trosze na chybi trafi. Trzeba szukać metod opartych na innej zasadzie. Powszechnie przyjętą jest metoda najmniejszych kwadratów. Opiera się ona na wymaganiu, by suma kwadratów odchyleń wariantów y od odnośnych wartości Y (względnie x od X) była możliwie mała. Na tej zasadzie za pomocą rachunku nieskończonościowego można wyprowadzić wzory dla współczynników równań regresji. Jeżeli, tak jak to było podane poprzednio, warianty zmiennej ewentualnej x

oznaczymy przez u, warianty zaś zmiennej y przez v, wzory te będą miały formę:

$$a_1 = \frac{\Sigma u^2 \Sigma v - \Sigma u \Sigma uv}{n \Sigma u^2 - (\Sigma u)^2}$$

$$b_1 = \frac{n \Sigma uv - \Sigma u \Sigma v}{n \Sigma u^2 - (\Sigma u)^2}$$

$$a_2 = \frac{\Sigma v^2 \Sigma u - \Sigma v \Sigma uv}{n \Sigma v^2 - (\Sigma v)^2}$$

$$b_2 = \frac{n \Sigma uv - \Sigma u \Sigma v}{n \Sigma v^2 - (\Sigma v)^2}$$

Równaniom regresji można nadać prostszą formę, w której nie będzie współczynników a_1 i a_2. Istotnie wzór dla pierwszego z tych współczynników można przekształcić, biorąc pod uwagę, że $\Sigma u = n\bar{u}$ i $\Sigma v = n\bar{v}$:

$$a_1 = \frac{n\bar{v} \Sigma n^2 - \Sigma u \Sigma uv}{n \Sigma u^2 - (\Sigma u)^2} = \frac{n\bar{v} \Sigma u^2 - \bar{v}(\Sigma u)^2 + \bar{v}(\Sigma u)^2 + \Sigma u \Sigma uv}{n \Sigma u^2 - (\Sigma u)^2} =$$

$$= \bar{v} + \frac{\bar{v}\,(\Sigma u)^2 - \Sigma u \Sigma uv}{n \Sigma u^2 - (\Sigma u)^2} = \bar{v} + \frac{\Sigma u\,(\bar{v} \Sigma u - \Sigma uv)}{n \Sigma u^2 - (\Sigma u)^2} =$$

$$= \bar{v} + \frac{n\bar{u}\,(\bar{v} \Sigma u - \Sigma uv)}{n \Sigma u^2 - (\Sigma u)^2} = \bar{v} - \frac{\bar{u}(n \Sigma uv - n\bar{v} \Sigma u)^2}{n \Sigma u^2 - (\Sigma u)} =$$

$$= \bar{v} - \bar{u} \frac{n \Sigma uv - \Sigma u \Sigma v}{n \Sigma u^2 - (\Sigma u)^2} = \bar{v} - b_1 \bar{u}$$

W ten sposób pierwsze równanie regresji przybiera prostszą formę z jednym tylko współczynnikiem b_1:

$$Y = \bar{v} + b_1 (x - \bar{u}). \qquad (1)$$

Podobnie drugie równanie otrzyma formę:

$$X = \bar{u} + b_2 (y - \bar{v}) \qquad (2)$$

Wzory na współczynniki b_1 i b_2 można także przekształcić, nadając im wygodniejszą do obliczeń formę. A więc dla pierwszego z nich będziemy mieli:

$$b_1 = \frac{n(\Sigma uv - \Sigma u \Sigma v)}{n\Sigma u^2 - (\Sigma u)^2} = \frac{n\Sigma uv - n\bar{u}\Sigma v}{n\Sigma u^2 - 2(\Sigma u)^2 + (\Sigma u)^2} =$$

$$= \frac{n\Sigma(u\bar{v} - \bar{u}v)}{n\Sigma u^2 - 2n\bar{u}\Sigma u + n^2\bar{u}^2} = \frac{\Sigma v(u - \bar{u})}{\Sigma u^2 - 2\bar{u}\Sigma u + n\bar{u}^2} =$$

$$= \frac{\Sigma v(u - \bar{u})}{\Sigma u^2 - 2\Sigma u\bar{u} + \Sigma \bar{u}^2} = \frac{\Sigma v(u - \bar{u})}{\Sigma(u - \bar{u})^2}$$

Prowadząc dalej te przekształcenia, można otrzymać inną jeszcze formę powyższego wzoru. Mianowicie od licznika jego można odjąć sumę $\bar{v}\Sigma(u - \bar{u})$, która jest równa zeru. Będziemy wtedy mieli:

$$b_1 = \frac{\Sigma v(u - \bar{u}) - \bar{v}\Sigma(u - \bar{u})}{\Sigma(u - \bar{u})^2} = \frac{\Sigma v(u - \bar{u}) - \bar{v}(u - \bar{u})}{\Sigma(u - \bar{u})^2} =$$

$$= \frac{\Sigma(u - \bar{u})(v - \bar{v})}{\Sigma(u - \bar{u})^2}$$

Podobnie dla drugiego współczynnika otrzymamy wzory:

$$b_2 = \frac{\Sigma u(v - \bar{v})}{\Sigma(v - \bar{v})^2} = \frac{\Sigma(u - \bar{u})(v - \bar{v})}{\Sigma(v - \bar{v})^2}$$

Współczynniki b_1 i b_2 noszą nazwę współczynników regresji.

Dla równań regresji formy (1) i (2) można dowieść, że bez względu na wartość spółczynników b_1 i b_2 suma dodatnich odchyleń wariantów u i zmiennej x od odnośnych wartości X jest równa sumie odchyleń ujemnych i tak samo suma dodatnich odchyleń wariantów v zmiennej y od odnośnych wartości Y jest równa sumie ujemnych.

Wyprowadzone w tym ustępie równania są zwykle pisane inaczej. Mianowicie zamiast symboli wariantów u i v populalacyj pisze się symbole odnośnych zmiennych ewentualnych x i y. Pod pewnym względem jest to wygodniejsze, trzeba tylko pamiętać o tym, że sumowanie dotyczy nie wartości wariantów, lecz samych wariantów. Zamiast średnich arytmetycznych $\bar{u}$ i $\bar{v}$ dla wariantów bierze się przy tym średnie zmiennych ewentualnych, ale nie zwykłe, lecz ważone, które oznacza się symbolami $\bar{\bar{x}}$ i $\bar{\bar{y}}$ a nie $\bar{x}$ i $\bar{y}$. Będziemy w ten sposób mieli równania formy następującej:

$$Y = \bar{\bar{y}} + b_1(x - \bar{\bar{x}}), \qquad X = \bar{\bar{x}} + b_2(y - \bar{\bar{y}})$$

$$b_1 = \frac{\Sigma\, y(x - \bar{\bar{x}})}{\Sigma(x - \bar{\bar{x}})^2} = \frac{\Sigma(x - \bar{\bar{x}})(y - \bar{\bar{y}})}{\Sigma(x - \bar{\bar{x}})^2}$$

$$b_2 = \frac{\Sigma\, x(y - \bar{\bar{y}})}{\Sigma(y - \bar{\bar{y}})^2} = \frac{\Sigma(x - \bar{\bar{x}})(y - \bar{\bar{y}}}{\Sigma(y - \bar{\bar{y}})}$$

To ogólnie przyjęte znakowanie będziemy stosowali w dalszym ciągu, mając zawsze na myśli powyższe zastrzeżenia.

Zastosujmy powyższe wzory do realnych przykładów. Weźmiemy z początku dla większej przejrzystości materiały złożone tylko z dziesięciu wariantów. Poszczególne wartości zmiennych ewentualnych będą tu na ogół reprezentowane przez pojedyncze warianty. Weźmiemy mianowicie (tab. 39,1) majowe średnie temperatury powietrza z 10-letniego okresu 1899—1908 dla miejscowości:

Kraków	50°04′ N	19°57′ E
Poznań	52°26′	16°56′
Kijów	50°27′	30°30′
Nerczinskij Zawod	51°19′	119°37′
Swerdłowsk (Jekaterinburg)	56°50′	60°38′
Tomsk	56°30′	84°58′

Tabela 39,1

Średnie majowe temperatury.

Data	Kraków	Poznań	Kijów	Nerczinskij Zawod	Swerdłowsk	Tomsk
1899	13.7	12.7	14 8	8 4	10,1	11.4
1900	12.3	12.0	14.7	10.2	12.1	12.1
1901	14.2	15.1	15.1	10.0	10.6	10.1
1902	10.3	10.5	12.9	5.7	9.8	7.1
1903	14.4	14.0	15.0	8 7	7 9	8.2
1904	12.4	12.2	12.4	9.9	11.4	12.2
1905	14.0	14.1	15.9	7.9	10.7	7.6
1906	14.8	15 3	18.4	11 0	13 0	7.2
1907	15.7	14.8	17.3	10.8	7 1	9.2
1908	15.1	14.7	14.7	8.4	8.3	12.7
Średnie	13 69	13.54	15.12	9.10	10.10	9.78

Zestawmy temperatury Krakowa z temperaturami pięciu innych miejscowości podanych w tabeli 39,1. Dla przykładu podam szczegółowe rachunki dla zestawienia Kraków — Kijów (tab. 39, 2). Temperatury Krakowa będziemy uważali ze zmienną x, temperatury Kijowa za y.

Tabela 39,2

Temperatury majowe Krakowa (x) i Kijowa (y).

Data	x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1899	13.7	+0.01	0.0001	14.8	−0.32	0.1024	− 0.0032
1900	12.3	−1 39	1 9321	14.7	−0.42	0 1764	+ 0 5838
1901	14 2	+0.51	0.2601	15.1	−0 02	0.0004	− 0.0102
1902	10.3	−3.39	11.4921	12.9	−2.22	4 9284	+ 7.5258
1903	14.4	+0.71	0 5041	15.0	−0.12	0.0144	− 0 0852
1904	12.4	−1 29	1 6641	12.4	−2.72	7.3984	+ 3 5088
1905	14 0	+0.31	0.0961	15.9	+0.78	0.6084	+ 0 2418
1906	14 8	+1.11	1 2321	18.4	+3.28	10.7584	+ 3 6408
1907	15.7	+2 01	4.0401	17,3	+2 18	4 7524	+ 4 3818
1908	15 1	+1.41	1.9881	14.7	−0 42	0.1764	− 0.5922
Sumy	136.9	+6.07 −6.07	23 2090	151.2	+6.24 −6.24	28.9 160	+19.8828 − 0.6908

Na podstawie tej tabeli znajdujemy $b_1 = 0.8269$, $b_2 = 0.6637$. W podobny sposób można obliczyć współczynniki regresji dla pozostałych pięciu zestawień (tab. 39,3).

Tabela 39,3

Współczynniki regresji dla średnich majowych temperatur dla różnych miejscowości w stosunku do Krakowa.

Miejscowości	b_1	b_2
Poznań	+0.9300	+0.9342
Kijów	+0.8269	+0.6637
Nerczinskij Zawod	+0.5731	+0.5709
Swerdłowsk	−0.4270	−0.3109
Tomsk	+0.0520	+0.0275

40. Charakterystyki korelacji. Z przykładów poprzedniego ustępu wynika, że charakter korelacji może być bardzo różny. Można wobec tego mówić o różnych jej stopniach i różnej ścisłości. Stopień korelacji podaje, o ile zmienia się — wzrasta lub zmniejsza się — zmienna Y skutkiem przyrostu zmiennej x o jedność względnie o ile zmienia się x, jeżeli zmienna y wzrasta o jedność. W jednym i w drugim wypadku miarą stopnia korelacji będą odnośne współczynniki regresji. Chcąc mieć jedną wspólną miarę trzeba wziąć średnią tych dwóch współczynników. Ponieważ te współczynniki są wielkościami różnego rodzaju, trzeba tu użyć średniej geometrycznej. Otrzymamy w ten sposób jako miarę stopnia korelacji nową funkcję, która nosi nazwę współczynnika korelacji.

Współczynnik korelacji jest zatem średnią geometryczną współczynników regresji.

$$r = \sqrt{b_1 b_2}$$

Nadaje mu się znak taki, jaki mają współczynniki regresji. Dla przykładów poprzedniego ustępu będzie on miał wartości:

Poznań	+ 0.932
Kijów	+ 0.741
Nerczinskij Zawod	+ 0.572
Swerdłowsk	— 0.364
Tomsk	+ 0.0378

Warto zaznajomić się bliżej z tą funkcją statystyczną. Wstawiając do równania $r = \sqrt{b_1 b_2}$ wzory współczynników regresji z poprzedniego ustępu, otrzymamy równanie:

$$r = \frac{\Sigma (x - \overline{\overline{x}})(y - \overline{\overline{y}})}{\sqrt{\Sigma (x - \overline{\overline{x}}) 2 \ (y - \overline{\overline{y}}) 2}}$$

Ponieważ $\Sigma (x - \overline{\overline{x}})^2 = n \sigma_x$ i $\Sigma (y - \overline{\overline{y}})^2 = n \sigma_y$, gdzie σ_x i σ_y są to średnie odchylenia zmiennych x i y, otrzymujemy wzór:

$$r = \frac{\Sigma (x - \overline{\overline{x}})(y - \overline{\overline{y}})}{n \sigma_x \sigma_y}$$

który może służyć do obliczenia omawianego współczynnika.

Następnie można zrobić jeszcze jedno przekształcenie:

$$r = \frac{\sum \frac{(x - \overline{\overline{x}})}{\sigma_x} \times \frac{(y - \overline{\overline{y}})}{\sigma_y}}{n}$$

Mamy teraz w równaniu znane już nam współrzędne normalne:

$$\xi = \frac{x - \overline{\overline{x}}}{\sigma_x}, \quad \eta = \frac{y - \overline{\overline{y}}}{\sigma_y}$$

Wobec tego można napisać:

$$r = \frac{\Sigma \xi \eta}{n}$$

Ten ostatni wzór dowodzi, że współczynnik korelacji jest pewnego rodzaju momentem drugiego rodzaju, podobnie jak wskaźnik skłonności drugiego rodzaju i wskaźnik skupienia wariantów (por. ust. 19 i 20). Wyróżnia się on od wszystkich momentów, z jakimi mieliśmy dotychczas do czynienia tym, że figurują w nim dwie zmienne: jest to moment mieszany.

Za pomocą współczynnika korelacji można wyprowadzić dla współczynników regresji nowe wzory, które mogą być użyteczne:

$$b_1 = r\frac{\sigma_y}{\sigma_x}, \quad b_2 = r\frac{\sigma_x}{\sigma_y}$$

Jak to już było wspomniane, współczynnik korelacji uważa się za dodatni, jeżeli współczynniki regresji są dodatnie, za ujemny w przeciwnym razie. Jego wartość absolutna jest zawarta w granicach od 0 do 1. W ust. 29 przykłady zostały dobrane umyślnie dla zobrazowania tych różnic. Można dowieść, że współczynnik korelacji nie może być większy od jedności. Dowód jest zawiły, dlatego też nie podaje się. Czytelnik znajdzie go w podręczniku Andersona (str. 275—6).

Przechodzimy teraz do określenia ścisłości korelacji. Chodzi tu o wielkość odchyleń wariantów zmiennej y od odnośnych wartości Y, względnie o podobne odchylenia x od X. Ponieważ te odchylenia są różne, trzeba wziąć ich średnią. Najbardziej poglądowa byłaby średnia arytmetyczna ich wartości absolutnych, czyli osobliwego rodzaju odchylenie przeciętne, ale na to nie można wyprowadzić ogólnego wzoru. Trzeba wobec tego wziąć pewnego rodzaju odchylenie średnie — pierwiastek kwadratowy z sumy kwadratów tych odchyleń podzielonej przez ich liczbę:

$$\sqrt{\frac{\Sigma(y-Y)^2}{n}} \quad \text{względnie} \quad \sqrt{\frac{\Sigma(x-X)^2}{n}}$$

Dla tych wielkości nie trudno jest wyprowadzić wzór ogólny. A więc dla pierwszej z nich będziemy mieli: $Y = \overline{\overline{y}} + b_1 (x - \overline{\overline{x}})$ i wobec tego

$$\sqrt{\frac{\Sigma (y - Y)^2}{n}} =$$

$$= \sqrt{\frac{\Sigma (y - \overline{\overline{y}})^2 - 2\, b_1\, \Sigma (x - \overline{\overline{x}}) + b_1^2\, \Sigma (x - \overline{\overline{x}})^2}{n}} =$$

$$= \sqrt{\sigma_y^2 + b_1^2 \sigma_x^2 - 2\, b_1 \frac{\Sigma (x - x)(y - y)}{n}} =$$

$$= \sqrt{\sigma_y^2 + b_1^2\, \sigma_x^2 - 2\, b_1 \frac{\Sigma (x - \overline{\overline{x}})(y - \overline{\overline{y}})}{\Sigma (x - \overline{\overline{x}})^2} \frac{\Sigma (x - \overline{\overline{x}})^2}{n}} =$$

$$= \sqrt{\sigma_y^2 + b_1^2\, \sigma_x^2 - 2\, b_1^2\, \sigma_x^2} = \sqrt{\sigma_y^2 - b_1^2\, \sigma_x^2}\,.$$

Ponieważ zaś według przytoczonego poprzednio wzoru $b^1 = \frac{\sigma_y}{\sigma_x}\, r$,otrzymamy ostatecznie:

$$\sqrt{\frac{\Sigma (y - Y)^2}{n}} = \sigma_y \sqrt{1 - r^2}$$

Podobnie wypadnie:

$$\sqrt{\frac{\Sigma (x - X)^2}{n}} = \sigma_x \sqrt{1 - r^2}$$

Dla oceny tych średnich odchyleń trzeba je porównać z wielkością wariantów. W tym celu trzeba je podzielić przez odnośne średnie. Otrzymamy w ten sposób wielkości:

$$\frac{\sigma_y}{\overline{\overline{y}}} \sqrt{1 - r^2} \text{ i } \frac{\sigma_x}{\overline{\overline{x}}} \sqrt{1 - r^2}$$

analogicznie do wskaźnika zmienności.

Dla charakterystyki korelacji trzeba mieć jedną taką wielkość wspólną dla obu populacyj. Podobnie jak przy współczynniku korelacji trzeba będzie wziąć średnią geometryczną:

$$\sqrt{\frac{\sigma_x \sigma_y}{\bar{\bar{x}}\,\bar{\bar{y}}}\left(1 - r^2\right)}$$

Otrzymana wielkość będzie charakteryzowała wymiar omawianych odchyleń. Jest ona zawsze mniejsza od jedności i tym większa im ścisłość korelacji jest mniejsza. Wobec tego dla charakteryzacji ścisłości korelacji lepiej jest wziąć dopełnienie do jedności:

$$\omega = 1 - \sqrt{\frac{\sigma_x \sigma_y}{\bar{\bar{x}}\,\bar{\bar{y}}}\left(1 - r^2\right)}$$

Tę wielkość można nazwać w s k a ź n i k i e m ś c i s ł o ś c i k o r e l a c j i. Wskaźnik ten jest mniejszy od jedności. Tylko przy $r = 1$ dochodzi on do tej granicy. Ścisłość korelacji jest wtedy największa i każdej wartości x odpowiada jedna tylko wartość y i odwrotnie.

Przykłady na wskaźnik ścisłości korelacji będą podane w następnych ustępach, bo dla przytoczonych powyżej materiałów klimatologicznych obliczyć go nie można, chyba że się wprowadzi temperatury absolutne.

41. Wiarogodność charakterystyk korelacji. Tak samo jak dla wszystkich innych charakterystyk, trzeba teraz rozpatrzyć wiarogodność charakterystyk korelacji. Trzeba zobaczyć, w jakim stopniu wartości tych charakterystyk otrzymane z prób różnią się od odnośnych wartości dla całości populacji. Trzeba wziąć, tak jak w ust. 28, wszystkie możliwe próby z porównywanych populacyj i dla każdej pary prób obliczyć wartości charakterystyk korelacji. Otrzymamy w ten sposób wtórne populacje, które dadzą możność rozwiązania postawionego w ten sposób zadania. Jest ono bardzo trudne i dlatego ograniczymy się do konkretnego przykładu.

Weźmiemy mianowicie 90 par danych, z których pierwsza jest liczbą promieni w szczytowych baldachach olszewnika (*Selinum carvifolia*), druga zaś liczbą promieni w pierwszym bocznym baldachu tejże rośliny (tab 41,1). Widzimy od razu, że baldachy szczytowe mają na ogół więcej promieni niż boczne. Liczby dla szczytowych baldachów będziemy uważali za warianty zmiennej ewentualnej x, liczby dla bocznych — za warianty zmiennej y.

Obliczmy dla tych dwóch populacyj jeden ze współczynników regresji oraz współczynnik korelacji. Wypadnie:

$$b_1 = +0.4321, \quad r = +0.679.$$

Tabela 41,1

Liczby promieni w szczytowych i pierwszych bocznych baldachach olszewnika.

31	33	27	32	24	25	28	26	28	29
23	29	27	25	23	23	26	21	27	21
40	41	34	24	34	30	27	32	33	29
27	24	26	22	28	25	25	29	25	24
30	40	33	29	28	28	31	25	24	28
24	27	30	18	18	22	23	18	18	24
26	29	40	25	24	29	32	27	26	29
24	21	28	21	17	25	26	26	22	23
36	27	24	31	38	31	24	32	29	30
23	23	19	29	29	26	21	26	29	22
33	28	29	25	23	26	31	34	25	28
25	25	26	23	23	21	25	28	23	26

22	27	33	24	27	22	39	29	29	28
21	24	25	22	23	25	30	25	24	26
39	33	38	37	29	28	29	28	26	26
30	30	30	28	23	22	25	22	20	26
35	29	26	32	28	27	29	30	28	21
26	24	27	25	23	22	21	25	23	21

Weźmiemy następnie z omawianego materiału 9 prób po 10 par wariantów według tabeli 41,1 i zobaczymy, jakie wartości wypadną dla tych samych współczynników. Wartości b_1, i r są zestawione w tabeli 41, 2. Są to częściowe populacje wtórne tych charakterystyk korelacji.

Tabela 41,2

Baldachy szczytowe i boczne olszewnika.

Próby	b_1	r
1	0.3926	0.432
2	0.1572	0.404
3	0.7007	0.744
4	0 5044	0.728
5	0.4776	0 602
6	0.4252	0 648
7	0 3908	0.834
8	0.6278	0.837
9	0 2898	0 519
Średnie	0.4407	0.639

Wypadły one różnie dla różnych prób i odmiennie dla prób niż dla całości materiału. Średnie arytmetyczne wartości dla prób zbliżają się natomiast do wartości odnośnych charakterystyk dla całości materiału. Zupełnego wyrównania nie

można się spodziewać nawet wtedy, gdyby się wzięło wszystkie możliwe próby. Mielibyśmy wtedy wprawdzie pełne populacje wtórne, ale jak to było wyjaśnione w ust. 29, tylko populacje wtórne złożone ze średnich dają średnią równą wartości charakterystyki dla populacji.

Zbadanie pełnych populacyj wtórnych dla charakterystyk korelacji prowadzi do określenia wymiaru średnich błędów dla tych charakterystyk. Otrzymuje się wzory:

$$\varepsilon(b_1) = \frac{\sigma_y \sqrt{1 - r^2}}{\sigma_x \sqrt{n - 2}}, \quad \varepsilon(b_2) = \frac{\sigma_x \sqrt{1 - r^2}}{\sigma_y \sqrt{n - 2}}, \quad \varepsilon(r) = \frac{1 - r^2}{\sqrt{n - 1}}$$

Dla podanych powyżej populacyj dla olszewnika mamy $\sigma_x = 4.465$, $\sigma_y = 2.840$ i wobec tego wypada $\varepsilon(b_1) = 0.1651$, $\varepsilon(r) = 0.245$.

Co do wzorów na średni błąd współczynników regresji i korelacji trzeba zrobić to samo zastrzeżenie, jakie było obszernie omówione odnośnie podobnych wzorów dla charakterystyk pojedynczych populacyj w ust. 30. Mianowicie jako tako dokładne wartości otrzymuje się z nich tylko przy dużej liczbie wariantów. Przy małej ich liczbie trzeba obliczać, tak jak dla średniej arytmetycznej, prawdopodobieństwo gorszych — bardziej odbiegających od prawdziwej — wartości niż ta, którą się otrzymało dla danej próby (por. ust. 31). Wykonanie takich obliczeń znajdzie czytelnik w podręcznikach Fishera i Romanowskiego.

Trzeba dodać jeszcze jedno zastrzeżenie: podane wzory dla błędów charakterystyk korelacji zostały wyprowadzone w założeniu, że rozpatrywane populacje mają formę normalną. Ponieważ tak na ogół nie jest, wynikają z tego dalsze odchylenia wartości obliczonych od prawdziwych.

Można naturalnie postawić zagadnienie wiarogodności wskaźnika ścisłości korelacji. Wobec zawiłości tego zagadnienia nie będę go rozpatrywał. Ograniczę się do podania wartości wspomnianego wskaźnika dla materiału olszewnika. Dla

niego mamy: $\sigma_x = 4.465$, $\sigma_y = 2.840$, $\bar{x} = 29.47$, $\bar{y} = 24.28$, $r = 0.679$. Wobec tego wypada $\omega = 0.902$.

42. Tabele korelacyjne. Przy dużych populacjach trzeba dla uproszczenia obliczeń układać dane w tabele korelacyjne, albowiem wtedy te same pary wartości x i y występują wielokrotnie. W tym celu rysuje się szachownicę i wypisuje się u góry pionowych rzędów wartości x a na początku poziomych wartości y. Następnie na skrzyżowaniu jednych i drugich wpisuje się częstości odnośnych par wartości x i y. Wreszcie u dołu pionowych rzędów umieszcza się częstości wartości zmiennej x, a na końcu poziomych rzędów częstości wartości zmiennej y.

Jako przykład podam: 1) liczby dzieci u matki i córki według danych Pearsona, Lee i Moore'a, 2) wiek męża i żony według angielskiego spisu ludności z r. 1901, 3) liczby kwiatów języczkowych i rurkowych w szczytowych koszykach arniki według badań własnych nad materiałem z Łojowej w okolicach Delatyna i 4) średnie zachmurzenie i dobowe amplitudy temperatury powietrza w Batawii na Jawie z lat 1908—1917. Pomijam natomiast klasyczny przykład korelacji między wzrostem syna i ojca, przytaczany nieomal we wszystkich podręcznikach, gdyż odnośne materiały angielskie są wadliwe z powodu niewłaściwego wytknięcia granic klas, przez co częstości wypadły częściowo ułamkowe.

Dla obliczenia charakterystyk korelacji z takich tablic metodą najmniejszych kwadratów używa się tych samych wzorów, jakie poprzednio stosowaliśmy do małych materiałów, a więc:

$$b_1 = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2}$$

$$b_2 = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (y - \bar{y})^2}$$

$$r = \sqrt{b_1 b_2}$$

Tabela 42,1

Korelacja między liczbą dzieci u matki i u jednej z jej córek w małżeństwach trwających conajmniej 15 lat według angielskiej statystyki.

Liczba dzieci u córki	Liczba dzieci u matki																Ogółem
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
0	5	9	11	18	21	15	8	9	6	3	2	3	—	—	—	—	110
1	12	5	14	15	10	13	9	8	5	3	2	2	—	—	—	—	98
2	9	9	10	15	18	15	9	3	2	4	2	—	—	—	—	1	97
3	5	10	16	11	9	14	13	10	4	8	2	3	—	—	—	—	105
4	5	5	19	17	21	15	18	10	14	2	1	5	1	—	—	—	133
5	7	6	7	17	23	9	12	13	14	8	3	2	2	—	—	—	123
6	4	5	8	11	15	12	15	14	7	5	3	3	1	—	—	—	103
7	5	4	3	8	4	13	9	8	5	10	2	1	1	—	—	—	73
8	1	2	4	12	9	9	8	5	12	3	4	1	2	1	—	—	73
9		—	4	3	3	4	7	5	3	2	2	1	—	—	—	—	34
10	—	—	1	2	1	3	4	6	3	2	—	1	—	1	—	—	24
11	—	—	2	1	1	1	—	—	1	2	—	—	—	—	—	—	8
12	—	2	1	2	3	—	1	1	—	—	1	—	1	—	1	—	13
13	—	—	—	—	2	1	—	—	—	—	1	—	2	—	—	—	6
Ogółem	53	57	100	132	140	124	113	92	76	52	25	22	10	2	1	1	1000

Tabela 42,2

Korelacja między wiekiem żony i męża według angielskiego spisu ludności z r. 1901. Częstości oznaczają tysiące par.

Wiek mężów (y)	Wiek żon (x)															Ogółem
	15—	20—	25—	30—	35—	40—	45—	50—	55—	60—	65—	70—	75—	80—	85—	
15 -	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	4
20—	16	173	46	4	1	—	—	—	—	—	—	—	—	—	-	240
25—	4	185	402	84	10	2	1	—	—	—	—	—	—	—	-	688
30 -	1	41	265	411	84	12	2	1	—	—	—	—	—	—	—	817
35—	—	9	69	251	369	80	12	1	1	—	—	—	—	—	—	793
40—	—	3	17	71	219	309	66	12	2	1	—	—	—	—	—	700
45—	—	1	6	20	66	178	252	59	10	2	1	—	—	—	-	595
50—	—	—	2	8	19	57	146	195	44	10	2	—	—	—	—	483
55—	—	—	1	3	8	18	46	110	141	35	6	1	—	—	—	369
60—	—	—	—	1	3	8	16	39	81	101	23	4	1	—	—	277
65—	—	—	—	1	1	3	6	11	26	53	58	13	2	1	—	175
70—	—	—	—	—	1	1	2	5	8	18	31	31	6	1	—	104
75—	—	—	—	—	—	1	1	2	3	5	10	14	12	2	—	50
80—	—	—	—	—	—	—	—	1	1	1	2	4	5	3	1	18
85—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	—	4
Ogółem	23	414	808	854	781	669	550	437	317	226	134	68	27	8	1	5317

Tabela 42,3

Korelacja między liczbą kwiatów języczkowych i rurkowych w szczytowych koszykach arniki z okolic Łojowej.

Liczby kwiatów rurkowych (y)	Liczby kwiatów języczkowych (x)																		Ogółem
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
43—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
48	—	—	2	-	1	—	—	—	—	—	—	—	—	—	—	—	—	—	3
53—	1	—	—	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	5
58—	—	—	1	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	3
63—	—	—	—	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	4
68—	—	—	—	—	1	—	1	1	2	—	—	—	—	—	—	—	—	—	5
73—	—	—	1	1	—	3	—	—	—	1	—	—	—	—	—	—	—	—	6
78—	—	—	1	1	2	—	4	—	1	2	1	1	—	—	—	—	—	—	13
83—	—	—	2	2	4	4	7	7	3	2	3	—	—	—	—	—	—	—	35
88—	—	—	1	—	3	9	12	4	7	6	1	—	—	1	—	—	—	—	44
93—	—	1	1	—	4	3	11	10	6	10	2	—	2	—	-	—	—	—	50
98—	—	—	1	1	3	5	13	14	15	11	6	2	1	—	—	—	—	—	72
103—	—	—	—	3	3	7	11	21	13	10	11	2	5	2	1	—	—	—	89
108—	—	—	1	1	1	4	7	12	7	8	13	5	4	5	—	—	—	—	68
113—	—	—	—	—	—	3	11	18	16	10	10	12	1	1	1	1	—	—	84
118—	—	—	1	—	1	2	4	16	13	21	12	9	5	4	1	—	—	—	89
123—	—	—	—	3	1	—	7	5	5	16	9	15	15	6	1	—	—	—	83
128 -	—	—	—	-	—	-	5	7	2	9	14	11	9	2	5	2	—	—	66
133 -	—	—	—	—	1	1	3	1	2	6	7	6	5	3	—	1	1	1	38
138—	—	—	—	—	—	1	1	1	1	8	4	8	4	4	3	2	—	1	38
143 -	—	—	—	—	—	—	2	1	3	6	6	3	9	4	1	2	—	—	37
148—	-	—	—	—	—	—	—	-	1	5	2	5	11	2	—	1	—	—	27
153—	—	—	—	—	—	—	—	—	—	2	—	1	3	1	4	1	1	1	14
158—	—	—	—	—	—	—	—	—	—	3	1	3	—	1	2	—	2	1	13
163—	—	—	—	—	—	—	—		—	—	1	—	4	—	1	2	—	—	8
168 -	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	2
173—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
178—	—	—	—	—	—	—	—	—	—	-	—	—	—	1	—	—	—	—	1
Ogółem	1	1	12	15	27	45	101	118	98	136	104	85	78	38	20	12	5	4	899

Tabela 42,4

Korelacja między średnim zachmurzeniem a dzienną amplitudą temperatury powietrza w Batawii w miesiącach czerwiec — sierpień 1908 — 1917 lat.

Dzienne amplitudy temperatury (y)	Średnie zachmurzenie w setnych (x)																			Ogółem
	5—	10—	15—	20—	25—	30—	35—	40—	45—	50—	55—	60—	65—	70—	75—	80—	85—	90—	95—	
2.0—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	2
2.5—	-	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3.0—	-	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	3	5
3.5—		—	—	—	—	-	—	—	—	—	—	—	2	—	—	—	2	3	2	9
4.0—	—	—	—	—	—	—	—	—	—	1	—	—	2	—	1	—	-	—	5	9
4.5 -	—	—	—	—	—	—	—	—	—	1	—	2	1	2	2	3	1	1	6	19
5.0—	—	—	—	—	—	-	—	—	—	1	5	1	3	2	6	6	4	4	2	34
5.5 -	—	—	—	-	—	—	—	1	6	7	1	5	2	7	3	5	6	3	2	48
6.0—	—	-	—	—	—	—	2	5	5	6	12	13	6	11	14	6	5	5	—	90
6.5—	—	—	2	3	2	1	8	12	9	17	15	14	16	6	5	5	5	1	1	122
7.0—	—	—	2	5	10	14	13	24	18	17	10	16	14	12	6	5	2	1	1	170
7.5—	—	1	5	8	16	14	27	20	20	16	9	13	6	7	2	1	2	1	—	168
8.0—	—	2	5	5	12	14	18	15	9	7	6	4	—	3	—	1	—	—	—	101
8.5 -	—	4	4	5	12	5	9	4	4	4	3	-	2	1	—	1	—	—	—	58
9.0—	—	3	5	4	9	8	7	6	3	3	2	—	-	—	—	—	—	—	—	50
9.5—	1	—	1	5	3	5	—	1	—	—	2	—	—	—	—	—	—	—	—	16
10.0—	1	—	4	2	1	3	—	—	—	1	—	—	—	-	—	—	—		—	12
10.5—	—	—	1	2	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	4
11.0	—	2	—	—	1	—	—	—	—	—	—	—	—	—	—	-	—	—	—	3
Ogółem	2	12	29	37	66	64	84	89	74	81	65	68	54	51	40	34	27	19	24	920

Przede wszystkim trzeba obliczyć sumę $\Sigma (x - \bar{x}) (y - \bar{y})$. W tym celu trzeba uciec się do metody wartości wyjściowej. Nietrudno jest wyprowadzić odnośny wzór. Przyjmujemy dla zmiennej x wartość wyjściową a_1, dla zmiennej y — wartość a_2. Będziemy mieli:

$$\Sigma (x - a_1)(y - a_2) = \Sigma (x - \bar{x} + \bar{x} - a_1)(y - \bar{y} + \bar{y} - a_2) =$$
$$= \Sigma (x - \bar{x})(y - \bar{y}) + \Sigma (x - \bar{x})(\bar{y} - a_2) + \Sigma (\bar{x} - a_1)(y - \bar{y}) +$$
$$+ \Sigma (\bar{x} - a_1)(\bar{y} - a_2) = \Sigma (x - \bar{x})(y - \bar{y}) + (\bar{y} - a_2)\Sigma (x - \bar{x}) +$$
$$+ (\bar{x} - a_1)\Sigma (y - \bar{y}) + n(\bar{x} - a_1)(\bar{y} - a_2).$$

Ponieważ sumy $\Sigma (x - \bar{x})$ i $\Sigma (y - \bar{y})$ jako sumy algebraiczne odchyleń od średniej są równe zeru, drugi i trzeci wyraz znikają. Otrzymujemy zatem:

$$\Sigma (x - \bar{x})(y - \bar{y}) = \Sigma (x - a_1)(y - a_2) - n(\bar{x} - a_1)(\bar{y} - a_2).$$

Ale z drugiej strony $\bar{x} = a_1 + \dfrac{\Sigma (x - a_1)}{n}$

$$\bar{y} = a_2 + \frac{\Sigma (y - a_2)}{n}$$

Wobec tego będziemy mieli ostateczny wzór:

$$\Sigma (x - \bar{x})(y - \bar{y}) = \Sigma (x - a_1)(y - a_2) - \frac{\Sigma (x - a_1)\Sigma (y - a_2)}{n}$$

Sumowanie dotyczy naturalnie wszystkich par wariantów obu zmiennych.

Pozostaje jeszcze obliczyć sumy kwadratów odchyleń od średniej. Jest to zadanie już nam znane, równoważne z obliczaniem średniego odchylenia. Dla zmiennej x mamy na to wzór:

$$\Sigma (x - \bar{x})^2 = \Sigma (x - a_1)^2 - \frac{[\Sigma (x - a_1)]^2}{n}$$

i podobny wzór dla drugiej zmiennej:

$$(y - \bar{y})^2 = (y - a_2)^2 - \frac{[\Sigma (y - a_2)]^2}{n}$$

W praktycznym wykonaniu porządek obliczeń jest odwrotny. Pokażę go na przykładzie liczb dzieci u matki i córki (tab. 42,1). Obliczenie sum kwadratów odchyleń, zestawione w tabelach 42,5 — 6 nie wymaga żadnych wyjaśnień.

Przyjmujemy $a_1 = 5$, $a_2 = 4$. Otrzymujemy $\Sigma (x - a_1) = +398$, $\Sigma (y - a_2) = +335$ oraz $\Sigma (x - a_1)^2 = 8814$, $\Sigma (y - a_2)^2 = 8919$.

Stąd $$\Sigma (x - \bar{\bar{x}})^2 = 8814 - \frac{898^2}{1000} = 8814 - 806 = 8008$$

$$\Sigma (y - \bar{\bar{y}})^2 = 8919 - \frac{335^2}{1000} = 8919 - 112 = 8807$$

Tabela 42,5

x	f	$x - a_1$	$f(x - a_1)$	$f(x - a_1)^2$
1	53	— 4	212	848
2	57	— 3	171	513
3	100	— 2	200	400
4	132	— 1	132	132
5	140	0	— 715	
6	124	+ 1	124	124
7	113	+ 2	226	452
8	92	+ 3	276	828
9	76	+ 4	304	1216
10	52	+ 5	260	1300
11	25	+ 6	150	900
12	22	+ 7	154	1078
13	10	+ 8	80	640
14	2	+ 9	18	162
15	1	+ 10	10	100
16	1	+ 11	11	121
	1000		+ 1613	8814

Tabela 42,6

y	f	$y - a_2$	$f(y - a_2)$	$f(y - a_2)^2$
0	110	— 4	440	1760
1	98	— 3	294	882
2	97	— 2	194	388
3	105	— 1	105	105
4	133	0	— 1033	
5	123	+ 1	123	123
6	103	+ 2	206	412
7	73	+ 3	219	657
8	73	+ 4	292	1168
9	34	+ 5	170	850
10	24	+ 6	144	864
11	8	+ 7	56	392
12	13	+ 8	104	832
13	6	+ 9	54	486
	1000		+ 1368	8919

Teraz przychodzi z kolei bardziej zawiłe obliczenie sumy $\Sigma (x - \bar{\bar{x}}) (y - \bar{\bar{y}})$. W tym celu piszemy na nowo tabelę korelacyjną pozostawiając w niej wolne rzędy, odpowiadające wartościom wyjściowym (tab. 42,7), przez co tabela zostaje podzielona na cztery „kwadraty". Następnie u góry piszemy odchylenia zmiennej x od wartości wyjściowej a_1, z lewej zaś strony takie same odchylenia y od a_2 i wreszcie w każdej kratce piszemy u góry z lewej strony odnośne częstości. Po tych wstępnych czynnościach przystępujemy do obliczeń. W każdej kratce u dołu z prawej strony piszemy iloczyn częstości przez odchylenie $y - a_2$. Te iloczyny sumujemy osobno dla każdego kwadratu i wpisujemy sumy w wolnym poziomym rzędzie. Sumy będą dodatnie w dolnych kwadratach, ujemne — w górnych.

Tabela 42,7

	− 4	− 3	− 2	− 1	0	+ 1	+ 2	+ 3	+ 4	+ 5	+ 6	+ 7	+ 8	+ 9	+ 10	+ 11
− 4	5 20	9 36	11 44	18 72		15 60	8 32	9 36	6 24	3 12	2 8	3 12				
− 3	12 36	5 15	14 42	15 45		13 39	9 27	8 24	5 15	3 9	2 6	2 6				
− 2	9 18	9 18	10 20	15 30		15 30	9 18	3 6	2 4	4 8	2 4					1 2
− 1	5 5	10 10	16 16	11 11		14 14	13 13	10 10	4 4	8 8	2 2	3 3				
0	− 79 + 34	− 79 + 52	−122 + 96	−158 +161		−143 +162	− 90 +168	− 76 +154	− 47 +131	− 37 + 96	− 20 + 58	− 21 + 26	+ 41	+ 10	+ 8	− 2
+ 1	7 7	6 6	7 7	17 17		9 9	12 12	13 13	14 14	8 8	3 3	2 2	2 2			
+ 2	4 8	5 10	8 16	11 22		12 24	15 30	14 28	7 14	5 10	3 6	3 6	1 2			
+ 3	5 15	4 12	3 9	8 24		13 39	9 27	8 24	5 15	10 30	2 6	1 3	1 3			
+ 4	1 4	2 8	4 16	12 48		9 36	8 32	5 20	12 48	3 12	4 16	1 4	2 8	1 4		
+ 5			4 20	3 15		4 20	7 35	5 25	3 15	2 10	2 10	1 5				
+ 6			1 6	2 12		3 18	4 24	6 36	3 18	2 12		1 6		1 6		
+ 7			2 14	1 7		1 7			1 7	2 14						
+ 8		2 16	1 8	2 16			1 8	1 8			1 8		1 8		1 8	
+ 9						1 9					1 9		2 18			

Tabela 42,8

Liczba dzieci u matki i córki.

$x-a_1$	(+)	(−)	$\Sigma=$ =(+) + (−)	$(x-a_1)\Sigma$
− 4	34	79	−45	+ 180
− 3	52	79	−27	+ 81
− 2	96	122	−26	+ 52
− 1	161	158	+ 3	− 3
0				
+ 1	162	143	+19	+ 19
+ 2	168	90	+78	+ 156
+ 3	154	76	+ 78	+ 234
+ 4	131	47	+84	+ 336
+ 5	96	37	+59	+ 295
+ 6	58	20	+38	+ 228
+ 7	26	21	+ 5	+ 35
+ 8	41	—	+41	+ 328
+ 9	10	—	+10	+ 90
+10	8	—	+ 8	+ 80
+11	—	2	− 2	− 22
				+2114
				− 25
				+2089

Otrzymane wyniki zbieramy w osobną tabelę 42,8. A mianowicie piszemy w pierwszej kolumnie odchylenia $x - a_1$, w osobnych kolumnach (drugiej i trzeciej) sumy dodatnie i ujemne iloczynów z poprzedniej tabeli. W następnej kolumnie (czwartej) umieszczamy sumy algebraiczne tych sum osobno dla każdego odchylenia $x - a_1$. Następnie mnożymy te ostatnie sumy przez wspomniane odchylenia, a iloczyny w ten sposób otrzymane piszemy w ostatniej kolumnie. Wreszcie sumujemy te iloczyny, uważając na znaki — dodawanie ma być algebraiczne! Otrzymujemy w ten sposób:

$$\Sigma (x - a_1)(y - a_2) = + 2089$$

W końcu odejmujemy poprawkę:

$$\frac{\Sigma(x-a_1)\,\Sigma(y-a_2)}{n}=\frac{898\times 335}{1000}=300.83,$$

co daje ostatecznie $\Sigma(x-\bar{x})(y-\bar{y})=2089-301=1788.$

Na podstawie otrzymanych w ten sposób wyników wypada ostatecznie:

$$b_1=\frac{1788}{8008}=+0.2233,\ b_2=\frac{1788}{8807}=+0.2030,\ r=+0.2129.$$

Równania regresji będą miały formę:

$$Y=4.33+0.2233\,(x-5.90)$$
$$X=5.90+0.2030\,(y-4.33)$$

Dla obliczenia średnich błędów będą jeszcze potrzebne wartości średnich odchyleń σ_x i σ_y. Wynoszą one w naszym przypadku $\sigma_x=2.830$, $\sigma_y=2.968$. Wobec tego $\varepsilon(b_1)=0.0324$, $\varepsilon(b_2)=0.095$, $\varepsilon(r)=0.0302$.

Wreszcie wskaźnik ścisłości korelacji jest równy $\omega=0.440$. Mamy w tym przypadku niewielką wartość współczynnika korelacji i mierną jej ścisłość.

Pozostałe przykłady potraktuję mniej szczegółowo, pozostawiając czytelnikowi trud wykonania obliczeń. A więc przykład drugi, dotyczący korelacji między wiekiem żony i męża, różni się od poprzedniego tym, że warianty są zgrupowane. Niewielka z tego wynika komplikacja, gdyż szerokość klas jest dla obu zmiennych jednakowa. mianowicie 5. Operując numerami klas, jak wartościami zmiennych ewentualnych. otrzymujemy wartości:

$$\Sigma(x'-\bar{x}')^2=34347,\quad (y'-\bar{y}')^2=36497.$$

Trzeba od nich odjąć poprawkę S h e p p a r d a w wysokości określonej wzorem $\frac{n}{12}$, wypadnie wtedy:

$$\Sigma(x'-\bar{x}')^2=33904,\quad (y'-\bar{y}')^2=36054.$$

Następnie suma $\Sigma(x'-\bar{x}')(y'-\bar{y}')$ będzie równa 32374.

Wobec tego, że szerokość klas jest jednakowa dla obu zmiennych, możemy otrzymane wartości użyć bez zmian do dalszych obliczeń, tak jak gdyby $x' = x$, $y' = y$ otrzymamy:

$$b_1 = \frac{32374}{33904} = 0.9550, \quad b_2 = \frac{32374}{36054} = 0.8980, \quad r = 0.926,$$

$$\varepsilon(b_1) = 0.0053, \quad \varepsilon(b_2) = 0.0050, \quad \varepsilon(r) = 0.052, \quad \omega = 0.883.$$

Równania regresji będą $Y = 42.31 + 0.955\,(x - 40.06)$
$X = 40.06 + 0.898\,(y - 42.31)$.

Mamy tu w przeciwieństwie do pierwszego przykładu wysokie wartości współczynnika korelacji i wskaźnika ścisłości.

Większe komplikacje będą w trzecim przykładzie, dotyczącym liczb kwiatów rurkowych i języczkowych w szczytowych koszykach arniki (tab. 42,3), gdyż liczby kwiatów rurkowych są zgrupowane, liczby zaś kwiatów języczkowych niezgrupowane. Szerokość klas dla kwiatów rurkowych jest 5. Oglądając tabelę korelacyjną, czytelnik zapewne zdziwi się, dlaczego zakresy klas nie są wytknięte przez liczby wielokrotne 5. Otóż zostały one ustalone w ten sposób, żeby wartości środkowe wypadły wielokrotne 5.

Dla liczby kwiatów języczkowych (zmienna x) otrzymujemy $\Sigma(x - \bar{x})^2 = 7009$. Operując numerami klas dla liczby kwiatów rurkowych (zmienna y), będziemy mieli $\Sigma(y' - \bar{y}')^2 = 16365$, a po zastosowaniu poprawki S h e p p a r d a w wymiarze $-\frac{n}{12} = -75$ otrzymamy 16290. Suma $\Sigma(y - \bar{y})^2$ będzie 25 razy większa.

Przy obliczaniu sumy $\Sigma(x - \bar{x})(y - \bar{y})$ będziemy dla zmiennej y operowali numerami klas. Otrzymamy $\Sigma(x - \bar{x})(y' - \bar{y}') = 6388$. Suma $\Sigma(x - \bar{x})(y - \bar{y})$ będzie 5 razy większa.

Stąd otrzymamy:

$$b_1 = \frac{6388 \times 5}{7009} = 4.557, \quad b_2 = \frac{6388 \times 5}{16290 \times 25} = 0.0784, \quad r = 0.598$$

Równania regresji będą: $Y = 115.1 + 4.557\,(x - 18.73)$
$X = 18.73 + 0.0784\,(y - 115.1)$.

Nadto mając ciągle na uwadze grupowanie zmiennej y, będziemy mieli średnie błędy $\varepsilon(b_1) = 0.218$, $\varepsilon(b_2) = 0.0033$, $\varepsilon(r) = 0.021$ i wskaźnik ścisłości korelacji $\omega = 0.863$.

Wreszcie w ostatnim przykładzie, dotyczącym zachmurzenia i dobowej amplitudy temperatury powietrza, mamy obie zmienne zgrupowane, ale o różnej szerokości klas: wynosi ona dla zachmurzenia 0.05, dla amplitudy temperatury 0.5 czyli 10 razy więcej.

Dla zachmurzenia (zmienna x) otrzymujemy $\Sigma(x' - \bar{x}')^2 = 15878$, po zastosowaniu poprawki S h e p p a r d a 15801. Dla otrzymania sumy $\Sigma(x - \bar{x})^2$ trzeba tę liczbę pomnożyć przez kwadrat szerokości klas, to znaczy przez 0.0025.

Dla amplitud temperatury (zmienna y) będziemy mieli $\Sigma(y' - \bar{y}')^2 = 6322$, a po zastosowaniu poprawki S h e p p a r d a 6245. Dla otrzymania sumy $\Sigma(y - \bar{y})^2$ trzeba pomnożyć przez $0.5^2 = 0.25$.

Suma $\Sigma(x' - \bar{x}')(y' - \bar{y}')$ wypada ujemna równa -6780. Dla otrzymania $\Sigma(x - \bar{x})(y - \bar{y})$ trzeba tę liczbę pomnożyć przez $0.05 \times 0.5 = 0.025$.

Współczynniki będą miały wartości:

$$b_1 = -\frac{6780 \times 0.025}{15801 \times 0.0025} = -4.291,$$

$$b_2 = -\frac{6780 \times 0.025}{6322 \times 0.25} = -0.1086,$$

$$r = -0.629.$$

Równania regresji będą: $Y = 7.251 - 4.291\,(x - 0.515)$

$$X = 0.515 - 0.1086\,(y - 7.251)$$

Wreszcie będziemy mieli średnie błędy $\varepsilon(b_1) = 0.1614$, $\varepsilon(b_2) = 0.0041$, $\varepsilon(r) = 0.020$ i wskaźnik ścisłości korelacji $\omega = 0.792$.

ROZDZIAŁ VIII.

KONTYNGENCJA.

43. Pojęcie kontyngencji. Kontyngencją nazywa się współzależność między populacjami, z których jedna przynajmniej ma charakter jakościowy.

Jeżeli populacja ma charakter jakościowy, to jej warianty dają się podzielić na dwie grupy na podstawie posiadania względnie braku takiej czy innej cechy. Na przykład o ile chodzi o barwę włosów i oczu, taką cechą będzie występowanie barwika. W ten sposób włosy czarne, szatyniaste i rude będą miały tę cechę, włosy zaś blond są jej pozbawione. Z drugiej strony oczy piwne i czarne będą miały, niebieskie zaś szare i zielone są jej pozbawione. Występowanie danej cechy oznaczamy jakąś literą, powiedzmy *A*, brak jej natomiast taką samą literą ale małą, a więc *a*.

Rozpatrzmy teraz dwie populacje jakościowe, pozostające ze sobą w związku z jakiejkolwiek przyczyny, na przykład wspomniane powyżęj barwy włosów i oczu u tych samych osób. Oznaczamy występowanie cechy pierwszej populacji przez *A*, cechy drugiej przez *B*, nadto brak pierwszej z tych cech przez *a*, brak drugiej przez *b*. Otóż ogólnie można się spodziewać, że część wariantów *A* będzie miała cechę drugiej populacji, część zaś będzie jej pozbawiona i to samo można przypuszczać o wariantach *a*. Wobec tego można będzie jedną i drugą populację podzielić na cztery części. Warianty pierwszej części będą miały cechy i pierwszej, i drugiej populacji, oznaczymy je *AB*. Warianty drugiej będą miały tylko cechę

pierwszej, oznaczymy je przez *Ab*. Trzecia część wariantów będzie miała cechę tylko drugiej populacji, będą to warianty *aB*. Ostatnia wreszcie będzie pozbawiona cech obu populacyj, będą to warianty *ab*. Liczebność tych grup wariantów można zestawić w tabelę podobną do tabeli korelacyjnej ale złożoną tylko z czterech kratek (tab. 43,1). Te liczebności oznaczymy symbolami wariantów ujętymi w nawias.

Tabela 43,1

Cechy	A	a	Ogółem
B	(AB)	(aB)	(AB) + (aB)
b	(Ab)	(ab)	(Ab) + (ab)
Ogółem	(AB) + (Ab)	(aB) + (aB)	n = (AB) + (aB) + + (Ab) + (ab)

Jako konkretny przykład weźmiemy współzależność między barwą włosów i oczu według badań A m m o n a w Badenii (tab. 43,2). Nazwijmy ciemnymi włosy i oczy posiadające wyraźny barwik, jasnymi nie mające go.

Tabela 43,2

Oczy	W ł o s y		Ogółem
	Ciemne	Jasne	
Ciemne	742	115	857
Jasne	3229	2714	5943
Ogółem	3971	2829	6800

Tego rodzaju tabele służą za podstawę do oceny charakteru kontyngencji, jeżeli jedna z populacyj ma charakter ilościowy, mają one inną formę — składają się z dwóch szeregów

kratek. Na przykład Pearson ogłosił następujące dane dotyczące związku między wiekiem dzieci a anemią (tab. 43,3).

Tabela 43,3

Barwa krwi	Wiek dzieci							Ogółem
	7	8	9	10	11	12	13	
Naturalna	34	29	27	32	31	35	41	229
Blada	17	17	23	19	21	21	17	135
Ogółem	51	46	50	51	52	56	58	364

Trzeba jeszcze dodać, że każdą populację ilościową można potraktować jako jakościową. Można bowiem warianty większe od pewnej granicznej wartości uważać jako posiadające pewną cechę a mniejsze jako nie mające jej. Taki podział jest stosowany na przykład w antropologii, kiedy się wyróżnia osobniki długogłowe od krótkogłowych.

W następnym ustępie zajmiemy się określeniem charakteru kontyngencji. Ograniczymy się do najprostszego przypadku, kiedy są porównywane dwie populacje, obie jakościowe.

44. Współczynnik kontyngencji. Badanie kontyngencji ma zupełnie inny charakter niż rozpatrzone w poprzednim rozdziale badanie charakteru korelacji. Mianowicie dotyczy ono częstości wariantów, podczas gdy tam chodziło o ich wartości. To też nie można tu mówić o stopniu ścisłości, tylko ogólnie o charakterze kontyngencji.

Jak zawsze w statystyce, chodzi i w tym przypadku o ujęcie liczbowe. Trzeba ustalić jakiś współczynnik kontyngencji, któryby określał charakter tego rodzaju współzależności między populacjami. Taki współczynnik powinien przede wszystkim czynić zadość warunkowi, aby się równał zeru, skoro między cechami rozpatrywanych populacyj niema żadnego związku, to znaczy jeżeli częstość względna A będzie taka

sama dla wariantów mających i niemających cechy B i odwrotnie. Korzystając z oznaczeń tabeli 43,1 poprzedniego ustępu, można ten warunek zapisać w formie proporcji:

$$\frac{(AB)}{(aB)} = \frac{(Ab)}{(ab)} \quad \text{albo} \quad \frac{(AB)}{(Ab)} = \frac{(aB)}{(ab)}$$

Stąd wynika, że wzór na współczynnik kontyngencji powinien zawierać czynnik $(AB)(ab) - (Ab)(aB)$.

Dalej ten współczynnik przybierać powinien wartość plus jeden przy najsilniejszym powiązaniu cech A i B. Najsilniejsze powiązanie będzie wtedy, jeżeli każdemu wariantowi A odpowiada wariant B i odwrotnie. Zajdzie to wtedy, kiedy $(aB) = (Ab) = 0$. Wreszcie współczynnik kontyngencji powinien przybierać wartość minus jeden, jeżeli cechy A i B wyłączają się wzajemnie, to znaczy, jeżeli żadnemu wariantowi z cechą A nie odpowiada wariant z cechą B i odwrotnie. Taki przypadek wystąpi, jeżeli $(AB) = 0$.

Powyższym trzem warunkom czyni zadość zaproponowany przez Y u l e'a współczynnik:

$$\varrho = \frac{(AB)(ab) - (Ab)(aB)}{(AB)(ab) + (Ab)(aB)}$$

Takie rozwiązanie zagadnienia nie jest jedyne. Proponowano różne inne formy podobnego współczynnika. W szczególności często można spotkać się z tzw. w s p ó ł c z y n n i k i e m k o r e l a c j i d l a c e c h n i e w y m i e r z a l n y c h, określonym przez wzór:

$$R = \frac{(AB)(ab) - (Ab)(aB)}{\sqrt{[(AB)+(aB)]\,[(AB)+(Ab)]\,[(aB)+(ab)]\,[(Ab)+(ab)]}}$$

Taka funkcja ma tyle wspólnego ze współczynnikiem korelacji, że da się wyprowadzić według prawideł obliczenia wspomnianego współczynnika, jeżeli się założy, że warianty A i B mają wartość jeden, a warianty a i b — wartość zero. Takie założenie oczywiście nie ma żadnego sensu. Powyższy „współczynnik korelacji ma ten sam licznik co współczynnik

kontyngencji ϱ, a więc staje się równym zeru jednocześnie z nim. W innych przypadkach daje wartości odmienne, na przykład dla przytoczonego powyżej materiału statystycznego dotyczącego barwy włosów i oczu wypada $\varrho = 0.689$, $R = 0.201$.

Przytoczę jeszcze jeden przykład: skuteczność działania surowicy Behringa przy zastosowaniu jej do dzieci chorych na płonicę (dyfteryt). Fibiger ogłosił w tym względzie ciekawe dane z obserwacyj wykonanych w Kopenhadze (tab. 44,1). Wypada tu $\varrho = 0.591$, $R = 0.160$. Widzimy, że różnice pomiędzy obu funkcjami mającymi charakteryzować kontyngencje są ogromne. Ciekawy to przykład względności metod statystycznych.

Tabela 44,1

Współzależność między stosowaniem surowicy Behringa do dzieci chorych na płonicę a wyzdrowieniem.

	Wyzdrowienie	Smierć	Ogółem
Z surowicą	231	8	239
Bez surowicy	215	29	244
Ogółem	446	37	483

Tabela A

$$\varphi_0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	0	1	2	3	4	5	6	7	8	9
0.0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0.4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2.0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0044	0033	0024	0017	0012	0009	0006	0004	0003	0002
4.0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

Tabela B

$$Q = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

Dziesięciotysięczne

x	0	1	2	3	4	5	6	7	8	9
— 3.0	0014	0010	0007	0005	0003	0002	0001	0001	0001	0001
— 2.9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
— 2.8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
— 2.7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0027
— 2.6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
— 2.5	0062	0060	0058	0057	0055	0054	0052	0051	0049	0048
— 2.4	0082	0080	0078	0076	0073	0071	0070	0068	0066	0064
— 2.3	0107	0104	0102	0099	0097	0094	0091	0089	0087	0084
— 2.2	0139	0136	0132	0129	0126	0122	0119	0116	0113	0110
— 2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
— 2.0	0228	0222	0217	0212	0207	0202	0197	0192	0186	0183
— 1.9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
— 1.8	0359	0351	0344	0336	0329	0321	0314	0307	0301	0294
— 1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
— 1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
— 1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
— 1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
— 1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
— 1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
— 1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
— 1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
— 0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
— 0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
— 0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
— 0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
— 0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
— 0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
— 0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
— 0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
— 0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
— 0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641

Tabela B (dokończenie)

$$Q = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

Dziesięciotysięczne

x	0	1	2	3	4	5	6	7	8	9
0.0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0.1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0.2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0.3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0.4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0.5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0.6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0.7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0.8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0.9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1.0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1.1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1.2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1.3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1.4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1.5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1.6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1.7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1.8	9641	9649	9656	9664	9671	9678	9886	9693	9699	9706
1.9	9713	9719	9726	9732	9738	9734	9750	9756	9761	9767
2.0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2.1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2.2	9861	9864	9868	9871	9874	9878	9881	9884	9887	9890
2.3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2.4	9918	9920	9922	9924	9927	9929	9930	9932	9934	9936
2.5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2.6	9953	9955	9956	9957	9958	9960	9961	9962	9963	9964
2.7	9965	9966	9967	9968	9968	9970	9971	9972	9973	9974
2.8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2.9	9981	9982	9982	9983	9984	9894	9985	9985	9986	9986
3.0	9986	9990	9993	9995	9997	9998	9999	9999	9999	9999

Tabela C

$$\varphi_3(x) = \frac{d^3\varphi_0}{dx^3} = -(x^3 + 3x)\varphi_0$$

Dziesięciotysięczne

x	0	1	2	3	4	5	6	7	8	9
0.0	+0000	+0120	+0239	+0359	+0478	+0597	+0716	+0834	+0952	+1070
0.1	+1187	+1303	+1419	+1534	+1648	+1762	+1874	+1986	+2097	+2207
0.2	+2315	+2422	+2529	+2634	+2737	+2840	+2941	+3040	+3138	+3235
0.3	+3330	+3423	+3514	+3604	+3693	+3779	+3864	+3947	+4027	+4106
0.4	+4184	+4259	+4332	+4403	+4472	+4539	+4603	+4666	+4726	+4785
0.5	+4841	+4895	+4946	+4996	+5043	+5088	+5131	+5171	+5209	+5245
0.6	+5278	+5309	+5338	+5365	+5389	+5411	+5431	+5448	+5463	+5476
0.7	+5486	+5495	+5501	+5504	+5506	+5505	+5502	+5497	+5490	+5481
0.8	+5469	+5456	+5440	+5423	+5403	+5381	+5358	+5332	+5305	+5276
0.9	+5245	+5212	+5177	+5140	+5102	+5062	+5021	+4978	+4933	+4887
1.0	+4839	+4790	+4740	+4688	+4635	+4580	+4524	+4467	+4409	+4350
1.1	+4290	+4228	+4166	+4102	+4038	+3973	+3907	+3840	+3772	+3704
1.2	+3635	+3566	+3495	+3425	+3354	+3282	+3210	+3138	+3065	+2992
1.3	+2918	+2845	+2771	+2697	+2623	+2549	+2475	+2402	+2328	+2254
1.4	+2180	+2106	+2033	+1960	+1887	+1815	+1742	+1670	+1599	+1528
1.5	+1457	+1387	+1317	+1248	+1179	+1111	+1044	+0977	+0911	+0846
1.6	+0781	+0717	+0654	+0591	+0529	+0468	+0408	+0349	+0290	+0233
1.7	+0176	+0120	+0065	+0011	−0042	−0094	−0146	−0196	−0245	−0294
1.8	−0341	−0387	−0433	−0477	−0521	−0563	−0605	−0645	−0685	−0723
1.9	−0760	−0797	−0832	−0867	−0900	−0933	−0964	−0994	−1024	−1052
2.0	−1080	−1106	−1132	−1156	−1180	−1203	−1225	−1245	−1265	−1284
2.1	−1302	−1320	−1336	−1351	−1365	−1380	−1393	−1405	−1416	−1426
2.2	−1436	−1445	−1453	−1460	−1467	−1473	−1478	−1483	−1486	−1490
2.3	−1492	−1494	−1495	−1496	−1496	−1495	−1494	−1492	−1490	−1487
2.4	−1483	−1480	−1475	−1470	−1465	−1459	−1453	−1446	−1439	−1432
2.5	−1424	−1416	−1407	−1399	−1389	−1380	−1370	−1360	−1349	−1339
2.6	−1323	−1317	−1305	−1294	−1232	−1270	−1258	−1245	−1233	−1220
2.7	−1207	−1194	−1181	−1168	−1154	−1141	−1127	−1114	−1100	−1086
2.8	−1073	−1059	−1045	−1031	−1017	−1003	−0989	−0976	−0962	−0948
2.9	−0934	−0920	−0906	−0892	−0879	−0865	−0852	−0838	−0824	−0811
3.0	−0798	−0669	−0552	−0449	−0359	−0283	−0219	−0168	−0127	−0095
4.0	−0070	−0051	−0036	−0026	−0018	−0012	−0008	−0006	−0004	−0002

Tabela D

$$\varphi_4(x) = \frac{d^4\varphi_0}{dx^4} = (x^4 - 6 \times 2 + 3)\,\varphi_0$$

Dziesięciotysięczne

x	0	1	2	3	4	5	6	7	8	9
0.0	+11968	+11965	+11956	+11941	+11920	+11894	+11861	+11822	+11777	+11727
0.1	+11671	+11609	+11541	+11468	+11388	+11304	+11214	+11118	+11017	+10911
0.2	+10799	+10682	+10560	+10434	+10302	+10165	+10024	+9878	+9727	+9572
0.3	+9413	+9250	+9082	+8910	+8735	+8556	+8378	+8186	+7996	+7803
0.4	+7607	+7408	+7206	+7001	+6793	+6583	+6371	+6156	+5940	+5721
0.5	+5501	+5279	+5056	+4831	+4605	+4378	+4150	+3921	+3691	+3461
0.6	+3231	+3000	+2770	+2539	+2309	+2078	+1849	+1620	+1391	+1164
0.7	+0937	+0712	+0487	+0265	+0043	−0176	−0394	−0611	−0825	−1037
0.8	−1247	−1454	−1660	−1862	−2063	−2260	−2455	−2646	−2835	−3021
0.9	−3203	−3383	−3559	−3731	−3901	−4066	−4228	−4387	−4541	−4692
1.0	−4839	−4983	−5122	−5257	−5389	−5516	−5639	−5758	−5873	−5984
1.1	−6091	−6193	−6292	−6386	−6476	−6561	−6642	−6720	−6792	−6861
1.2	−6925	−6986	−7042	−7093	−7141	−7185	−7224	−7259	−7292	−7318
1.3	−7341	−7361	−7376	−7388	−7395	−7399	−7400	−7396	−7389	−7378
1.4	−7364	−7347	−7326	−7301	−7274	−7243	−7209	−7172	−7132	−7088
1.5	−7043	−6994	−6942	−6888	−6831	−6772	−6710	−6646	−6580	−6511
1.6	−6441	−6368	−6293	−6216	−6137	−6057	5975	−5891	−5806	−5720
1.7	−5632	−5542	−5452	−5360	−5267	−5173	−5079	−4983	−4886	−4789
1.8	−4692	−4593	−4494	−4395	−4295	−4195	−4095	−3995	−3894	−3793
1.9	−3693	−3592	−3492	−3392	−3292	−3192	−3093	−2994	−2895	−2797
2.0	−2700	−2603	−2506	−2411	−2316	−2222	−2129	−2036	−1945	−1854
2.1	−1765	−1676	−1588	−1502	−1416	−1332	−1249	−1167	−1086	−1006
2.2	−0927	−0840	−0774	−0700	−0626	−0554	−0483	−0414	−0346	−0279
2.3	−0214	−0150	−0088	−0027	−0033	+0092	+0148	+0204	+0258	+0311
2.4	+0362	+0412	+0461	+0508	+0554	+0598	+0641	+0683	+0723	+0762
2.5	+0800	+0836	+0871	+0905	+0937	+0968	+0998	+1027	+1054	+1080
2.6	+1105	+1129	+1152	+1173	+1194	+1213	+1231	+1248	+1264	+1279
2.7	+1293	+1306	+1317	+1328	+1338	+1347	+1355	+1363	+1369	+1375
2.8	+1379	+1383	+1386	+1387	+1390	+1391	+1391	+1391	+1389	+1388
2.9	+1385	+1382	+1378	+1374	+1369	+1364	+1358	+1351	+1345	+1337
3.0	+1330	+1231	+1107	+0969	+0829	+0694	+0570	+0460	+0365	+0284
4.0	+0218	+0165	+0123	+0090	+0065	+0047	+0033	+0023	+0016	+0010

Ust. 32
(70)

Tabela E

Funkcja $S(t)$ Studenta (Metron Vol. 5 N 3, 1925)

Tysięczne

t	$n=2$	3	4	5	6	7	8	9	10	11
0.0	500	500	500	500	500	500	500	500	500	500
1	532	535	537	537	538	538	538	539	539	539
2	563	570	573	574	575	576	576	577	577	577
3	593	604	608	610	612	613	614	614	614	615
4	621	636	642	645	647	648	650	650	651	651
0.5	648	667	674	678	681	683	684	685	686	686
6	672	695	705	710	713	715	716	717	718	719
7	694	722	733	739	742	745	747	748	749	750
8	715	746	759	766	770	773	775	777	778	779
9	733	768	783	790	795	799	801	803	804	805
1.0	750	789	804	813	818	822	825	827	828	830
1	765	807	824	834	839	843	846	848	850	851
2	779	824	842	852	858	862	865	868	870	871
3	791	838	858	868	875	879	883	885	887	889
4	803	852	872	883	890	894	898	900	902	904
1.5	813	864	885	896	903	908	911	914	916	918
6	822	875	896	908	915	920	923	926	928	930
7	831	884	906	918	925	930	934	936	938	940
8	839	893	915	927	934	939	943	945	947	949
9	846	901	923	935	942	947	950	953	955	957
2.0	852	908	930	942	949	954	957	960	962	963
2	864	921	942	954	960	965	968	970	972	974
4	874	931	952	963	969	973	976	978	980	981
6	883	938	960	970	976	980	982	984	986	987
8	891	946	966	976	981	984	987	988	990	991
3.0	898	952	971	980	985	988	990	992	992	993
2	904	957	975	984	988	991	992	994	995	995
4	909	962	979	986	990	993	994	995	996	997
6	914	965	982	989	992	994	996	996	997	998
8	918	969	984	990	994	996	997	997	998	998
4.0	922	971	986	992	995	996	997	998	998	999
2	926	974	988	993	996	997	998	998	999	999
4	929	976	989	994	996	998	998	999	999	999
6	932	978	990	995	997	998	999	999	999	1000
8	935	980	991	996	998	998	999	999	1000	
5.0	937	981	992	996	998	999	999	1000		
2	940	982	993	997	998	999	999	1000		
4	942	984	994	997	998	999	1000			
6	944	985	994	998	999	999	1000			
8	946	986	995	998	999	999	1000			
6.0	947	987	995	998	999	1000				

Tabela F (dokończenie)

Funkcja *S(t)* Studenta (Metron Vol. 5 N3, 1925)

Tysięczne

t	n=12	13	14	15	16	17	18	19	20	21
0.0	500	500	500	500	500	500	500	500	500	500
0.1	539	539	539	539	539	539	539	539	539	540
0.2	577	578	578	578	578	578	578	578	578	579
0.3	615	615	616	616	616	616	616	616	616	618
0.4	652	652	652	652	653	653	653	653	653	655
0.5	686	687	687	688	688	688	688	688	688	691
0.6	720	720	721	721	721	722	722	722	722	726
0.7	751	751	752	752	753	753	753	754	754	758
0.8	780	780	781	782	782	782	783	783	783	788
0.9	806	807	808	808	809	809	810	810	810	816
1.0	831	832	832	833	833	834	834	835	835	841
1.1	853	854	854	855	856	856	857	857	857	864
1.2	872	873	874	875	876	876	877	877	877	885
1.3	890	891	892	893	893	894	894	895	895	903
1.4	906	907	908	908	909	910	910	911	911	919
1.5	919	920	921	922	923	924	924	924	924	933
1.6	931	932	933	934	935	935	936	936	936	945
1.7	941	943	944	944	945	946	946	947	947	955
1.8	950	952	952	953	954	955	955	956	956	964
1.9	958	959	960	961	962	962	963	963	963	971
2.0	965	967	967	967	968	969	969	970	970	977
2.2	975	976	977	977	978	979	979	979	979	986
2.4	982	983	984	985	985	986	986	986	986	992
2.6	988	988	989	990	990	990	991	991	991	995
2.8	991	992	992	993	993	994	994	994	994	997
3.0	994	994	995	995	996	996	996	996	996	999
3.2	996	996	996	997	997	997	997	998	998	999
3.4	997	997	998	998	998	998	998	998	998	1000
3.6	998	998	998	999	999	999	999	999	999	1000
3.8	998	999	999	999	999	999	999	999	999	1000
4.0	999	999	999	999	999	1000	1000	1000	1000	1000
4.2	999	999	1000	1000	1000	1000	1000	1000	1000	1000
4.4	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

INSTYTUT BADAWCZY LEŚNICTWA

INSTITUT POLONAIS DES RECHERCHES FORESTIERES

Prace z serii D

(Podręczniki)

Nr 1. **Dr Feliks Jezierski** — Pozyskiwanie żywicy sosnowej z drzew żywych — 1938, str. 146, rys. 92 (wyczerpane).

Nr 2. **Dr Stanisław Tyszkiewicz** — Nasiennictwo leśne — (w druku).

Nr 3. **Inż. Henryk Orłoś** — Grzyby jadalne i trujące — (w druku).

Nr 4. **Dr Dezydery Szymkiewicz** — Zadania i metody statystyki.

Nr 5. **Dr Tadeusz Gieruszyński** — Dendrometria — (w druku).

